TAMPEREEN TEKNILLINEN YLIOPISTO
TAMPERE UNIVERSITY OF TECHNOLOGY

TIINA RISTILÄ

DIABETIC RETINOPATHY DETECTION WITH TEXTURE FEATURES

Master of Science Thesis

Examiners: Prof. Joni Kämäräinen and Prof. Hannu Eskola
Examiners and topic approved by the Faculty Council of the Faculty of Computing and Electrical Engineering on January 4, 2017

ABSTRACT

TIINA RISTILÄ: Diabetic retinopathy detection with texture features
Tampere University of Technology
Master of Science Thesis, 53 pages, 3 Appendix pages
August 2017
Master's Degree Programme in Electrical Engineering
Major: Biomeasurements and Bioimaging
Examiners: Professor Joni Kämäräinen and Professor Hannu Eskola

Keywords: computer-aided detection, diabetic retinopathy, microaneurysm, texture features, machine vision

Diabetic retinopathy is one of the leading causes of visual impairment and blindness in the world and the prevalence keeps increasing. It is a vascular disorder of the retina and a symptom of diabetes mellitus. The health of the retina is studied with non-invasive retinal imaging. However, the analysis of the retinal images is laborious, subjective and the number of images to be reviewed is increasing.

In this master's thesis, a computer-aided detection system for diabetic retinopathy, microaneurysms and small hemorrhages was designed and implemented. The purpose of this study was to find out, are texture features able to produce descriptive and efficient information for the retinal image classification and is the implemented system accurate. The process included image preprocessing, extraction of 21 texture features, feature selection and classification with a support vector machine. The retinal image datasets that were used for the testing were Messidor, DIARETDB1 and e-ophta.

The texture features were not successful when classifying the retinal images into diabetic retinopathy or normal. The best average accuracy was 69 % with 72 % average sensitivity and 66 % average specificity. The texture features are not that descriptive as global features with a whole retinal image. Additionally, the varying size of the images and variation within a class affected the classification. The second experiment studied the classification of images into microaneurysm or normal by dividing the retinal images into blocks. The texture features were successful when the images were divided into small blocks of size 50*50. The best average accuracy was 96 % with 96 % average sensitivity and 96 % average specificity. The texture features are more descriptive in the local setting since then they can extract finer details.

To ease the clinical workflow of ophthalmologists and other experts, the computer-aided detection system can lower the manual labor and make retinal image analysis more efficient, accurate and precise. To develop the systems further, an optic disc and image quality detectors are needed.

TIIVISTELMÄ

TIINA RISTILÄ: Diabeettisen retinopatian havaitseminen tekstuuripiirteillä
Tampereen teknillinen yliopisto
Diplomityö, 53 sivua, 3 liitesivua
Elokuu 2017
Sähkötekniikan diplomi-insinöörin tutkinto-ohjelma
Pääaine: Biomittaukset ja -kuvantaminen
Tarkastajat: professori Joni Kämäräinen ja professori Hannu Eskola

Avainsanat: tietokoneavusteinen havaitseminen, diabeettinen retinopatia, mikroaneurysma, konenäkö, tekstuuripiirteet

Diabeettinen retinopatia on silmän verkkokalvon verisuoniston sairaus, joka voi kehittyä diabetesta sairastavalle henkilölle. Se on yksi maailman merkittävimmistä syistä näön heikkenemiseen sekä sokeutumiseen ja sen esiintyvyys kasvaa jatkuvasti. Verkkokalvon terveyttä voidaan tutkia ja seurata ei-invasiivisesti silmänpohjakuvauksen avulla. Silmänpohjakuvien analysoiminen on kuitenkin työlästä sekä subjektiivista ja analysoitavien kuvien määrä on kasvussa.

Tässä diplomityössä suunniteltiin ja toteutettiin ohjelmisto, jonka tarkoitus on havaita diabeettinen retinopatia, mikroaneurysmat sekä pienet verenvuodot silmänpohjakuvista. Työn tarkoitus oli selvittää, pystyvätkö tekstuuripiirteet tuottamaan deskriptiivistä ja tehokasta tietoa silmänpohjakuvien luokittelua varten ja kuinka tarkka toteutettu ohjelmisto on. Ohjelmiston prosesseihin kuuluivat kuvan esikäsittely, 21 tekstuuripiirteen irrotus sekä piirteiden valinta ja luokittelu tukivektorikoneen (engl. *support vector machine*) avulla. Ohjelmistoa testattiin Messidor, DIARETDB1 ja e-ophta silmänpohjakuva-aineistoilla.

Kun silmänpohjakuvia luokiteltiin terveiksi tai diabeettiseksi retinopatiaksi, tekstuuripiirteet eivät onnistuneet tuottamaan deskriptiivistä tietoa. Paras keskimääräinen tarkkuus oli 69 %, keskimääräinen sensitiivisyys oli 72 % ja keskimääräinen spesifisyys oli 66 %. Tekstuuripiirteet eivät onnistu tuottamaan deskriptiivistä tietoa, kun niitä käytetään kokonaiseen silmänpohjakuvaan. Lisäksi kuvien vaihteleva koko ja luokan sisäinen vaihtelevuus vaikuttivat luokittelun tuloksiin. Toinen koe jakoi kuvat pienemmiksi osioiksi ja tutki kuvaosioiden luokittelua mikroaneurysmaksi tai normaaliksi. Tekstuuripiirteillä saatiin hyviä tuloksia, kun pienten osioiden koko oli 50*50. Paras keskimääräinen tarkkuus oli 96 %, keskimääräinen sensitiivisyys oli 96 % ja keskimääräinen spesifisyys oli 96 %. Tekstuuripiirteet ovat hyvin deskriptiivisiä pienemmällä pinta-alalla, koska silloin ne kykenevät paremmin kuvaamaan pienempiä yksityiskohtia.

Tämä konenäköön perustuva ohjelmisto, jonka tarkoituksena on tukea silmälääkäreitä ja muita asiantuntijoita diabeettisen retinopatian sekä mikroaneurysmien havaitsemisessa, voi vähentää manuaalista työtä ja tehdä silmänpohjakuvien analysoinnista tehokkaampaa, tarkempaa ja täsmällisempää. Ohjelmistoa voi tulevaisuudessa kehittää lisäämällä siihen automaattisen näköhermon ja kuvanlaadun tunnistuksen.

PREFACE

Life is a funny thing. I lost my way for a while and almost didn't see this day to come. However, I am really glad for this moment: writing this preface.

I want to thank Joni Kämäräinen for his help and encouragement through the whole thesis process. His positive attitude and genuine interest gave me great momentum. Also, thanks to Hannu Eskola for examining my thesis.

I want to thank Riku Hietaniemi. Working in Oulu was very nice experience and I learned a lot by working with Riku.

My dear friends Elli, Jaana and Onerva - your friendship and support mean a great deal to me. I want to thank my grandma Kaarina for my intellectual and cultural upbringing.

My fiancé Antti - I truly admire his patience during my thesis process. I thank him for the help with my thesis and for his support and love.

In Tampere, Finland, on July 27, 2017

Tiina Ristilä

CONTENTS

LIST OF FIGURES

LIST OF SYMBOLS AND ABBREVIATIONS

2D	Two-dimensional
3D	Three-dimensional
AUC	Area under curve
CLAHE	Contrast-limited adaptive histogram equalization
CT	Computed Tomography
DD	Disc diameter
FN	False negative
FOV	Field of View
FP	False positive
GLCM	Gray level co-occurrence matrix
IRMA	Intraretinal microvascular abnormalities
kNN	k-nearest neighbors
LPB	Local binary pattern
MRI	Magnetic Resonance Imaging
NICE	National Institute for Health and Clinical Excellence of UK
NPRD	Nonproliferative diabetic retinopathy
PDR	Proliferative diabetic retinopathy
ROC	Retinopathy Online Challenge
ROC curve	Receiver Operating Characteristic curve
SVM	Support vector machine
TN	True negative
TP	True positive

A	The amount of gray levels in an image
B	Blue component in RGB color space
C	Cyan component in CMY color space
$\boldsymbol{CM}$	A gray level co-occurrence matrix
D	The amount of rows in an image
E	The amount of columns in an image
E_5	Edge, Laws' vector mask
η	The spatial aspect ratio (minor) for a Gabor filter
f	Frequency of a sinusoid in a Gabor filter
f_{max}	Maximum frequency of a sinusoid in a Gabor filter
G	Green component in RGB color space
γ	The spatial aspect ratio (major) for a Gabor filter
$\boldsymbol{GF}$	The two dimensional Gabor filter
H	Hue component in HSV color space
h	A histogram
I	An image
i	A gray level value
j	A gray level value
L_5	Level, Laws' vector mask
M	Magenta component in CMY color space
μ	Mean
μ_3	Skewness
μ_4	Kurtosis
$\boldsymbol{p}$	The normalized GLCM
R	Red component in RGB color space

R_N	The amount of neighboring pixels
R_5	Ripple, Laws' vector mask
S	Saturation component in HSV color space
S_5	Spot, Laws' vector mask
σ	Standard deviation
σ^2	Variance
θ	The rotation of the plane wave and the Gaussian in a Gabor filter
u	The scale in a Gabor filter
V	Value component in HSV color space
v	The orientation in a Gabor filter
W_5	Wave, Laws' vector mask
x	x coordinate
Δx	Displacement in rows
Y	Yellow component in CMY color space
y	y coordinate
Δy	Displacement in columns

1. INTRODUCTION

Diabetes is a chronic metabolic disorder that is characterized by high blood sugar (high glucose). Insulin is the hormone that controls the amount of glucose and inadequate secretion from pancreas causes serious health problems and complications. World Health Organization has estimated that in 2014, there were globally 422 million adults who had diabetes diagnosed and the prevalence just keeps increasing (World Health Organization 2016). This number lacks the number of children with diabetes. It is estimated that 5–10 % of diabetes cases are type 1, where the pancreas does not produce enough insulin and approximately 90 % are type 2 cases, where the ability of the body to respond to insulin has suffered (Leu & Zonszein 2010).

Diabetic retinopathy is a complication of diabetes and it is a vascular disease of the retina (Figure 1b). It is caused by the long-term effects of high glucose levels on the microvascular system. Diabetic retinopathy is one of the considerable causes of visual impairment and blindness globally (World Health Organization 2016). The early detection of diabetic retinopathy with proper guidance can lower the risk of severe vision loss by 90 % (Garg & Davis 2009). Sometimes the patient is not even aware of the fact that he has diabetes especially in the case of type 2 diabetes.

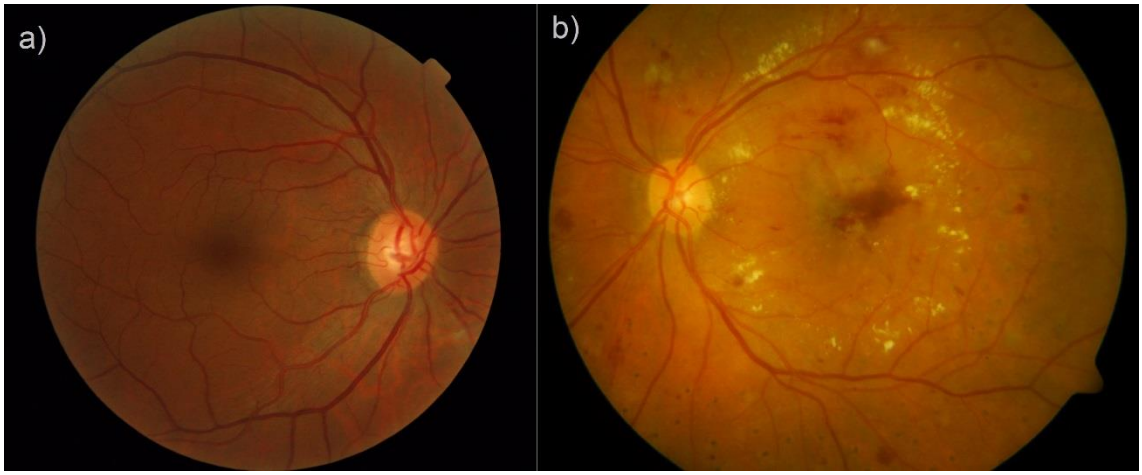

Figure 1. (a) A healthy retina and (b) a retina with the signs of diabetic retinopathy

To detect, diagnose and follow the development of diabetic retinopathy and other ocular diseases, non-invasive retinal imaging can be used. Retinal imaging is an optic imaging method. Ophthalmologists and other qualified medical experts analyze the retinal images. The process of analysis is subjective, time consuming and laborious. While the prevalence of diabetes and obesity (that is linked to type 2 diabetes) keep increasing, also the number

of retinal images that need analysis is increasing. The answer to these challenges could be found in medical image analysis and computer-aided detection.

Medical image analysis aims to classify various tissue types based on the question at hand and it has been applied for example to magnetic resonance imaging (MRI), computed tomography (CT) and X-ray imaging. Computer-aided detection systems help medical experts to detect tumors or lesions and to do that, these systems can use artificial intelligence and computer vision with image processing methods. For example, medical image analysis and computer vision have been applied to the quantitative analysis of the MRI images of the brain and thigh muscles (Sikiö 2016), the assessment of breast cancer in MRI images (Ahmed et al. 2013), the classification of gastroenterology images (Riaz et al. 2012) and the automated detection of tuberculosis in chest X-rays (Jaeger et al. 2014).

The aim of this thesis was to design and implement a computer-aided detection system for diabetic retinopathy and lesions. The purpose was to study whether texture features are able to produce descriptive and efficient information for the retinal image classification. In addition, how accurate the implemented system would be for the detection of a diabetic retinopathy and related lesions.

The implemented system had high accuracy, sensitivity and specificity in red lesion detection. The selected texture features were able to produce efficient and descriptive information on a local setting. This system could help ophthalmologists and other experts in their clinical workflow by lowering manual labor and making the analysis more accurate, precise and efficient. In addition, this could enable an early detection of diabetic retinopathy and therefore improve the quality of life of diabetics.

This thesis was part of a TEKES funded project at University of Oulu. The aim of the project was to map out research possibilities by combining data analysis and ophthalmic health. Researcher Riku Hietaniemi from University of Oulu was guiding the represented implementation.

The thesis is divided into six chapters. Chapter 2 describes the basics of diabetes, diabetic retinopathy and retinal imaging. Chapter 3 introduces medical image analysis. First the subject is issued generally and then related to the retinal images. Chapter 4 presents the information about the methods used in this thesis and Chapter 5 describes the experiments and the results with discussion. Chapter 6 contains the conclusion of this thesis. Appendix A includes the equations of the gray level co-occurrence matrix related statistical parameters.

2. DIABETES AND RETINAL IMAGING

The retina is an interior tissue layer in the eye. When incoming light travels to the eye, the retina enables the conversion of light into nerve impulses and these impulses are converted into images in our brains. Diseases and complications, such as diabetes or cardiovascular diseases, affect the health of the retina. Diabetic retinopathy is a complication of diabetes mellitus and it is one of the considerable causes of visual impairment and blindness globally (Pascolini & Mariotti 2012). Retinal imaging is used to analyze ocular structures, evaluate the health of the retina and to detect and diagnose possible diseases such as diabetic retinopathy. Retinal images are also used in follow-up to diseases and treatments.

2.1 Diabetes

Diabetes mellitus is a metabolic disorder that is characterized by high blood sugar (hyperglycemia). To control high glucose levels in blood stream, beta cells in pancreatic islets produce and secrete insulin hormone. Insulin increases the uptake and utilization of glucose into cells and as a result the concentration of glucose in blood stream starts to decline into normal (70–110 mg/dl). If the insulin production is inadequate, the blood glucose levels stay high and excess glucose is released into urine. On the other hand, body can become resistant to insulin if its concentration in blood stream is continuously high. (Martini 2006)

Diabetes has been classified into four main categories: type 1, type 2, gestational and other. Type 1 diabetes is an autoimmune disease of the pancreas and it usually advances quite slowly. In the end, it leads into insulin deficiency and Type 1 diabetic is dependent on insulin in a form of a replacement hormone. The type 2 diabetes is caused by insulin resistance and dysfunctional insulin secretion. It can be hard for a person to notice that they have the symptoms of type 2 diabetes (Table 1) since the disease is linked with other medical problems such as central obesity, high blood pressure and unbalanced cholesterol levels. These are significant cardiovascular risk factors. Management of the type 2 diabetes includes healthy lifestyle and medication. Gestational diabetes develops during pregnancy and the last category includes less common types of diabetes. (Leu & Zonszein 2010)

Table 1. Symptoms of type 1 and type 2 diabetes

Type 1 diabetes	Type 2 diabetes
• Frequent urination • Sudden weight loss • Constant thirst and dry mouth • Blurred vision • Fatigue • Constant hunger	• Frequent urination • Weight loss • Excessive thirst • Blurred vision

There is a strong link between cardiovascular diseases and diabetes. Vascular tissues suffer from high levels of glucose in the blood stream due to the resulting biochemical reactions. These reactions affect the inner cells and muscle cells of vessels. One major pathway to vessel damage is glycosylation. When glucose is circulating in blood stream, it can unintentionally attach to protein molecules and affect their normal functioning. Managing the blood glucose levels is challenging. Cardiovascular problems and microvascular complication arise from poorly managed blood glucose control. They lead for example into foot ulcers and amputation, kidney failure and vision loss. Cardiovascular disease is the most probable cause of death of type 2 diabetics. (Devereux 2010; Laud & Shabto 2010; Leu & Zonszein 2010)

2.2 Retina

The retina is the inner layer of the eye (Figure 2). The retina itself has two layers called the pigmented part and the neural part. The neural part senses the light that enters the eye and the pigmented part absorbs it. The light creates biochemical reactions in retina's photoreceptors and then, nerve impulses are sent through the optic nerve and visual pathway into the visual cortex of the brain. This part of the brain then forms images i.e. our vision. (Martini 2006)

The major blood vessels of the retina are radiating from the center of the optic nerve. The beginning of the optic nerve in the retina is called the optic disc. This part is also known as the blind spot since there are no photoreceptors. The spot with the highest concentration of cone photoreceptors is called fovea, the area of clear vision. The cone photoreceptors are responsible for giving clear vision in full colors. Macula (or macula lutea) is the area around the fovea. (Martini 2006)

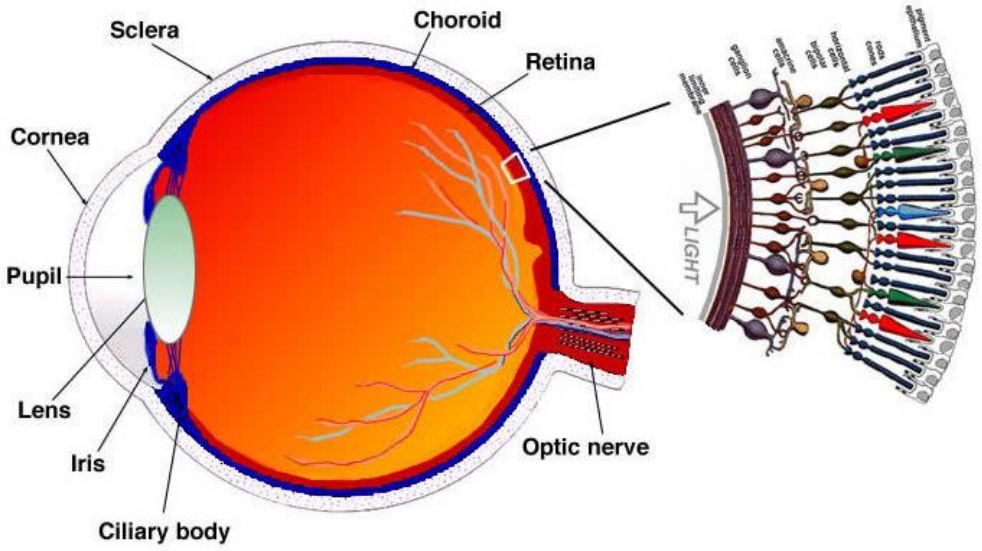

Figure 2. Anatomy of the retina (Kolb 1995)

2.3 Diabetic retinopathy

A person with diabetes mellitus can develop diabetic retinopathy, which is a vascular disorder of the retina. Probability to develop diabetic retinopathy gradually increases over the years. It is easier to calculate a time domain for type 1 diabetic since the onset of type 2 diabetes is usually quite uncertain. Another risk factor is blood glucose control, which affects capillaries. Regular follow-ups are important. (Laud & Shabto 2010)

The rupture, degeneration and excessive growth of blood vessels invade the space between the neural layer and pigmented layer. Visual acuity is threatened by a leakage of blood, new blood vessels and damage of photoreceptors. There is no cure but healthy lifestyle can help to prevent the progress and temporary fixes such as laser therapy are an option. (Martini 2006; Laud & Shabto 2010)

Microaneurysms are the earliest clinical signs of diabetic retinopathy (Figure 3a). Microaneurysms are located in retinal capillaries in any part of the retina and they appear as round and small red dots. The number of microaneurysms can predict the progression of retinopathy if they are increasing, decreasing or if the amount stays the same. The increasing number of microaneurysms indicate the progression of the disease. *Hemorrhages* have variety in their shape and size (Figure 3b). Very small hemorrhages can be hard to distinguish from microaneurysms while larger hemorrhages can resemble blots. Hemorrhages form when microaneurysm or weak vascular wall breaks within the retina. Microaneurysms and hemorrhages are also referred to as red lesions. (Chen 2008; Laud & Shabto 2010)

Exudates are also connected to diabetic retinopathy. *Hard exudates* are leaked lipoproteins and appear as white or yellow (Figure 3c). These exudates have strict outlines. *Soft exudates* are called cotton wool spots due to their soft shape (Figure 3d). Soft exudates arise from the nerve fiber layer where axoplasm is leaking. Hard exudates and cotton wool spots are also referred to as bright lesions. (Chen 2008)

When diabetic retinopathy progresses into a more severe state, the signs of abnormal growth of blood vessels start to surface. *Intraretinal microvascular abnormalities* (IRMA) form loops of vessels that splay other vessels. *Neovascularization* can lead into leaking vessels, sudden vision loss or even retinal detachment (Figure 3e). In venous beading, a vessel has both thinning and dilating parts. (Laud & Shabto 2010)

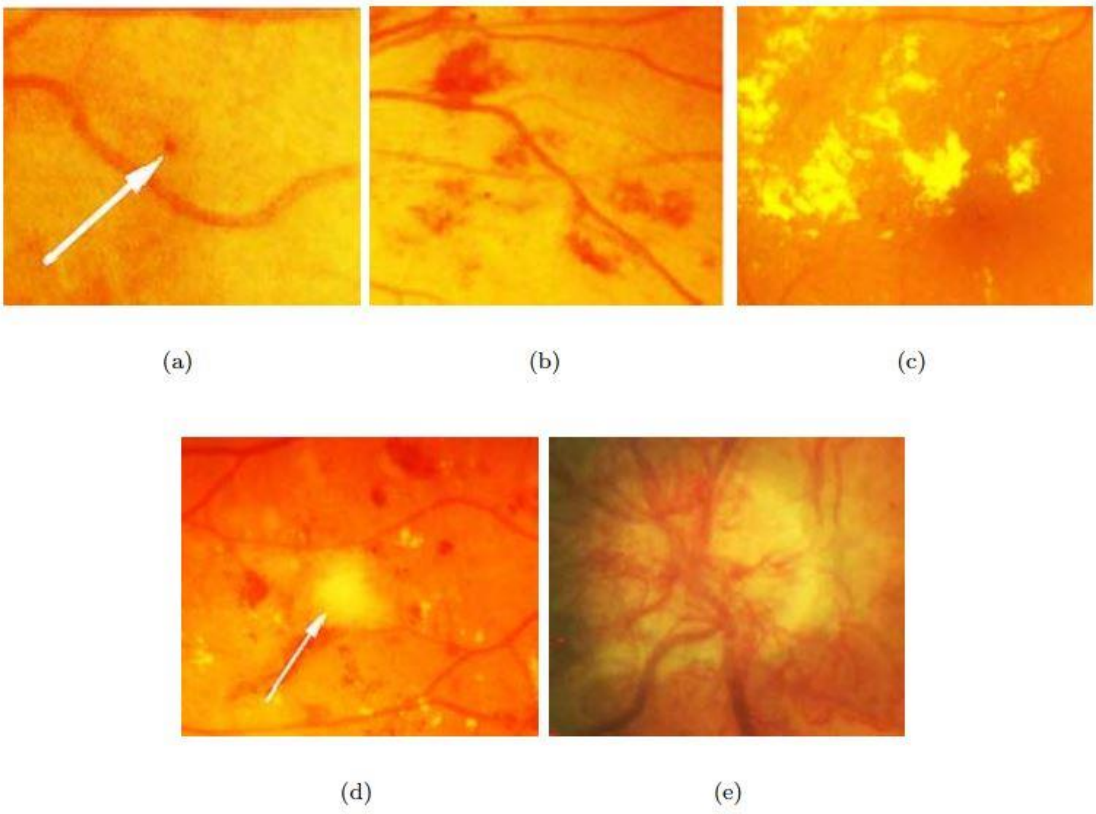

Figure 3. Symptoms of diabetic retinopathy. (a) A microaneurysm, (b) hemorrhages, (c) hard exudates, (d) soft exudate and (e) neovascularization in the optic disc. (Kauppi 2010)

The abnormalities in the retina can define two categories: nonproliferative diabetic retinopathy (NPDR) or proliferative diabetic retinopathy (PDR). NPDR is also known as background retinopathy. Lesions that are associated with NPDR are microaneurysms, hemorrhages, cotton wool spots and hard exudates. NPDR has three stages and it can advance into PDR with the abnormal growth of blood vessels. Table 2 represents connections between the findings and severity of retinopathy. (Laud & Shabto 2010)

Table 2. Diabetic Retinopathy Disease Severity Scale (Wilkinson et al. 2003)

Proposed Disease Severity Scale	Findings Observable on Dilated Ophthalmoscopy
No apparent retinopathy	No abnormalities
Mild nonproliferative diabetic retinopathy	Microaneurysms only
Moderate nonproliferative diabetic retinopathy	More than just microaneurysms but less than severe nonproliferative diabetic retinopathy
Severe nonproliferative diabetic retinopathy	Any of the following; more than 20 intraretinal hemorrhages in each of 4 quadrants; definite venous beading in 2+ quadrants; prominent intraretinal microvascular abnormalities (IRMA) in 1+ quadrant AND no signs of proliferative retinopathy
Proliferative diabetic retinopathy (PDR)	One or more of the following: neovascularization, vitreous/preretinal hemorrhage

Diabetic macular edema is sight threatening condition where retinal lesions are located near the center of the macula. In the severe stage, hard exudates have reached the center. Macular edema can include thickening of the retina or hard exudates. In NPDR, the macular edema is the primary reason for visual loss. (Wilkinson et al. 2003; Chen 2008; Laud & Shabto 2010)

2.4 Retinal image acquisition

Fundus photography is an optical method and therefore it is based on reflected light, which projects the retinal tissues to a two-dimensional (2D) representation. Fundus imaging is used to detect for example age-related degeneration and diabetic retinopathy. An expert makes the evaluation of the retina based on subjective analysis of colors, structures and shapes. (Abràmoff et al. 2010; Kaschke et al. 2013)

Fundus photography has a wide category of different modes. A typical mode is color fundus photography (Figure 4a), which is performed with white light. By filtering the light, user can highlight specific retinal structures. When a green filter is used, retinal blood vessels have higher contrast in a red-free image. Fluorescein angiography is used to detect retinal vessels, blood circulation and leakages (Figure 4c and 4d). First, sodium fluorescein dye is injected into a vein and then a series of images is obtained to detect the

circulation of the dye. Unfortunately, the dye can cause side effects such as nausea, vomiting, rash or even severe reactions. Stereo imaging mode produces a pair of monocular images that provide a stereoscopic view of for example optic nerve. (Abràmoff et al. 2010; Laud & Shabto 2010; Kaschke et al. 2013)

Optical coherence tomography (OCT) provides information about layered retinal structures and its histological changes (Figure 4b). It is based on a measurement of distance and time of flight of backscattering light. The wavelength of the light used is usually longer than visible light's wavelength. To get a 2D image, the beam of light moves in depth (z-axis) across the retina in transverse (x-axis) direction. The three-dimensional (3D) image is formed by stacking the 2D images at orthogonal positions (y-axis). The thickness and volume of the retina can be measured with OCT and therefore it is optimal to detect macular edema, macular degeneration, glaucoma and inflammatory diseases. (Abràmoff et al. 2010; Kaschke et al. 2013)

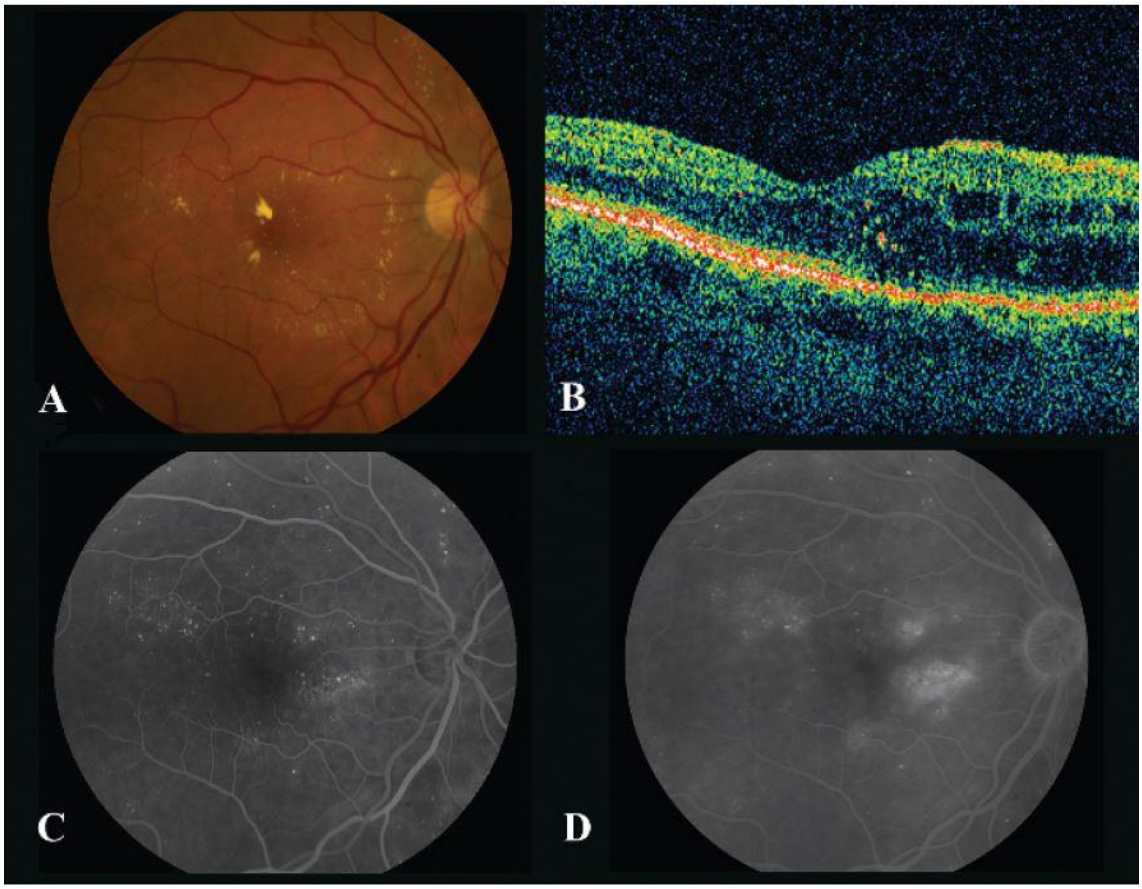

Figure 4. Retinal images. (A) Color fundus photography, (B) OCT: Optical coherence tomography and (C–D) Fluorescein angiogram (Meyerle et al. 2008)

2.5 Quality of retinal images

The homogeneity and good quality of retinal images decreases complexity and eases the analysis for humans and for computers. However, retinal imaging is technically challenging operation and images can include artefacts and wide technical variation due to users, patients or devices. In the worst case, the image cannot be evaluated.

There are several challenges in the retinal imaging. The smallest objects are in micrometers. Access to the retina is challenging since the light travels through a small hole i.e. through the pupil. The patient has to stay still since even a small shift in position affects the result. Some objects and structures have low contrast and need to be enhanced for example with fluorescein dye. When a 3D object is photographed and 2D image is produced, information gets lost. The curved shape of the fundus is also a challenge for the imaging. (Kaschke et al. 2013)

Artefacts can significantly affect the image quality. Figure 5 represents retinal images with poor quality. Typical artefacts are smudges on the lens, light artefact and non-uniform illumination. Small pupils, eye diseases and certain conditions that increase opacity can block or limit the light that enters the eye, which can therefore lower the sharpness of the image and create non-uniform illumination. This can create blurred or dark images. The color of the retina has variation due to age or ethnicity, which can create challenges to the assessment of the image. (Dias et al. 2014)

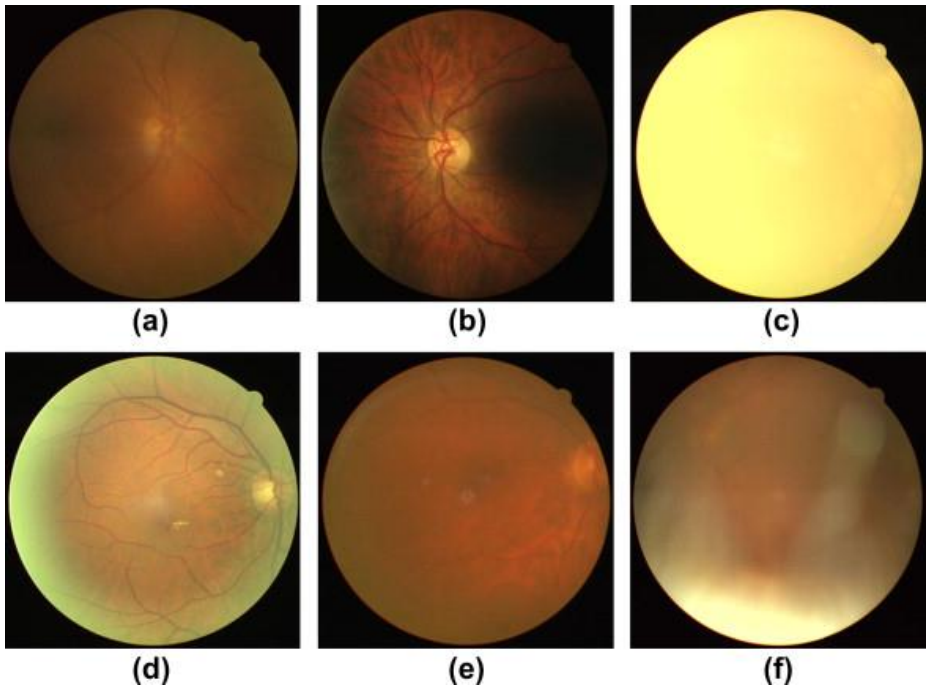

Figure 5. Retinal images with poor quality. (a) Poor focus, (b) Uneven illumination, (c) Blinked eye, (d) Light artefact, (e) Dust and dirt (near the center) and (f) eyelash artefact. (Dias et al. 2014)

Retinal imaging instruments from different manufactures have different settings and characteristics. The devices can have different Field of Views (FOV) and resolutions. Typical FOV is 50 degrees. Depending on the user of the device and purpose of the image, the images can have different views of focus (macula or optic disc). (Kaschke et al. 2013)

An ideal retinal image is free of reflections, artefacts and is uniformly illuminated. The image has high sharpness, FOV, contrast, resolution and the details are clearly visible. Table 3 describes informal quality specifications for technically satisfactory images from macula and optic disc view. The disc diameter (DD) is the diameter of the optic disc. The positioning of the image is important for the detection of abnormalities. (Taylor & Batey 2012; Kaschke et al. 2013)

Table 3. Definitions for technically satisfactory image (Taylor & Batey 2012)

View	Good	Adequate
Macula	The center of the fovea should be less than one *disc diameter* (1 DD) from the center of the image	The center of the fovea should be more than 2 DD from the edge of the image
	The vessels must be clearly visible within 1 DD of the center of the fovea	The vessels have to be visible within 1 DD of the center of the fovea
	The vessels ought to be visible across more than 90 % of the image	
Optic disc	The center of the disc has to be less than 1 DD from the center of the image	The complete optic disc must be more than 2 DD from the edge of the image
	The fine vessels should be clearly visible on the surface of the disc	The fine vessels are required to be visible on the surface of the disc
	The vessels are required to be visible across more than 90 % of the time	

2.6 Diabetic retinopathy datasets

There are several retinal image databases available online related to diabetic retinopathy. Seven image sets that are related to the detection of abnormalities or classification of diabetic retinopathy are introduced in the following subchapters.

2.6.1 Messidor

Messidor is an abbreviation of "Methods for Evaluating Segmentation and Indexing techniques Dedicated to Retinal Ophthalmology". Messidor dataset has 1 200 images where 546 images are classified as healthy and 645 images are classified as retinopathy. The images were acquired in three ophthalmologic departments with a Topcon TRC NW6 fundus camera before the year 2007. The field of view is 45 degrees and image sizes are 1440*960, 2240*1488 and 2304*1536 pixels. (MESSIDOR 2016)

The classification is divided into healthy (as class 0) and three stages of retinopathy severity (classes 1–3) based on microaneurysm, hemorrhage and neovascularization findings (Table 4). The images are classified based on retinal lesions but the locations and exact amounts of lesions are not reported. In addition, a risk factor for macular edema is reported with three classes. Class 0 represents absent risk for macular edema and classes 1–2 are at risk. The ground truths are provided as an Excel file. (MESSIDOR 2016)

Table 4. Retinopathy grade in Messidor database based on findings (MESSIDOR 2016)

Retinopathy grade	Findings
Class 0	No microaneurysms AND no hemorrhages
Class 1	Amount of microaneurysms is 1–5 AND no hemorrhages
Class 2	Amount of microaneurysms is 6–15 OR amount of hemorrhages is 1–4 AND no neovascularization
Class 3	Amount of microaneurysms is greater than 15 OR amount of hemorrhages is greater than 4 OR neovascularization is present

2.6.2 DIARETDB

As a subproject for the Imageret project, Kauppi et al. (2007) created two standard diabetic retinopathy datasets called DIARETDB0 and DIARETDB1. DIARETDB0 dataset has 130 images where 20 images have no signs of retinopathy and 110 images have at least mild NPDR signs. DIARETDB1 dataset has 89 images where 5 images have no signs of retinopathy and 84 images have at least mild NPDR signs. (Kauppi et al. 2007)

All the images were acquired in Kuopio University hospital before the year 2007. The DIARETDB1 set is acquired with one fundus camera while DIARETDB0 set has images by several fundus cameras. The field of view is 50 degrees and image size is 1500*1152 pixels. (Kauppi et al. 2007)

Four experts have annotated the images and marked hard and soft exudates, microaneurysms and hemorrhages. They were instructed to use shapes of centroid, polygon, circle and ellipse with a representative point in the middle of the shape. The experts also reported their confidence level considering the findings. The severity of diabetic retinopathy is not reported. The annotations are provided as XML files that include types, coordinates, confidence levels and shapes. (Kauppi et al. 2007)

2.6.3 Retinopathy Online Challenge (ROC)

Meindert Niemeijer, Michael D. Abramoff and Bram Van Ginneken organized the Retinopathy Online Challenge (ROC), which was an international competition for microaneurysm detection. They provided color fundus images divided as 50 training images and 50 test images. The training set included reference standard of microaneurysms and dot hemorrhages. The competition started February 2008. At the moment, the submissions are not possible but the files are available by download. (Niemeijer et al. 2010)

The images were selected from a larger set of 150 000 images and were acquired in different places. The images were acquired with either a Canon CR5-45NM, a Topcon NW 100 or a Topcon NW 200 camera. There are two image shapes present due to the different types of cameras. The coverage of the retina is 45 degrees and image sizes are 769*576, 1058*1061 and 1389*1383 pixels. (Niemeijer et al. 2010)

Four experts have annotated the images. They were asked to mark the center locations of microaneurysms and other lesions that look similar such as hemorrhages and pigment spots. Only the microaneurysms were included in the reference standard but in the competition, detecting similar lesions were not counted as a false positive. The annotations are provided as a XML file that includes coordinates and radii. (Niemeijer et al. 2010)

2.6.4 E-ophta

The e-ophta is a color fundus image database that is extracted from the Ophdiat project. Ophdiat was a five-year telemedicine screening program for diabetic retinopathy in France. The Ophdiat project started in 2004 and in total 51 741 examinations were made. (Schulze-Döbold et al. 2012; Decencière et al. 2013)

The e-ophta database is divided into e-ophta EX and e-ophta MA subsets. The e-ophta EX includes 35 healthy images and 12 278 annotated exudates within 47 images. The e-ophta MA set includes 233 images free of microaneurysms and 1 306 annotated microaneurysms and other small red lesion within 148 images. (Decencière et al. 2013)

The e-ophta images were acquired between the years of 2008 and 2009 in different places either with a Canon CR-DGi or with a Topcon TRC-NW6 camera. The coverage of the retina is 45 degrees and image sizes are 1440*960, 2048*1360 and 2544*1696 pixels.

The annotations are provided in a form of binary masks corresponding to the positions of microaneurysms or exudates in the original images. (Schulze-Döbold et al. 2012; Decencière et al. 2013)

2.6.5 Kaggle competition

Kaggle website runs machine learning related programming contest. Kaggle had a machine learning competition for diabetic retinopathy detection in 2015. The competition was popular since 661 teams and 854 contestants participated in it. The competition dataset had almost 90 000 images. The dataset has lots of variation since its purpose is to mimic real situation. (Kaggle 2015)

The training set included 35 126 images with ground truth, the test set included 53 576 images and both eyes of a patient were presented. The stage of diabetic retinopathy was marked by an expert. As presented on the Table 5, there is a scale from zero to four: no signs of diabetic retinopathy (class 0), mild (class 1), moderate (class 2), severe (class 3) and proliferative diabetic retinopathy (class 4). (Kaggle 2015)

Table 5. Kaggle training set images (Kaggle 2015)

Class	Name	Number of images	Percentage
0	No retinopathy	25 810	73.48 %
1	Mild NPDR	2 443	6.96 %
2	Moderate NPDR	5 292	15.07 %
3	Severe NPDR	873	2.48 %
4	PDR	708	2.01 %

The image acquiring methods, devices and conditions have lots of variation. Image sizes vary, but the images are on average 3000*2000 pixels, which is quite large. The images and labels include noise in a form of artefacts, underexposure, overexposure, out of focus images and outliers. (Kaggle 2015)

2.6.6 Diabetic Retinopathy Image Database (DRiDB)

Diabetic Retinopathy Image Database (DRiDB) was created by the image processing group in University of Zagreb, Croatia. The database contains 50 color fundus images. The images were acquired at the University Hospital of Zagreb with Zeiss Visucam 200 camera. The coverage of the retina is 45 degrees and image size is 720*576 pixels. The

set has 36 images that contain diabetic retinopathy signs and 14 images that do not contain any signs. (Prentašić et al. 2013)

The dataset contains annotations from several experts and at least five experts have annotated one image. The experts marked all fundus structures, pathologies and grade of the disease. They were asked to mark hard exudates, soft exudates, microaneurysms, hemorrhages, optic disc, macula, blood vessels and neovascularization. (Prentašić et al. 2013)

2.6.7 High resolution fundus (HRF)

The high resolution fundus (HRF) image database was a result of a collaborative research project by the University of Erlangen-Nuremberg, Germany and Brno University of Technology, Czech Republic. The image database includes 15 images of diabetic retinopathy and 15 images of healthy eyes. Diabetic retinopathy images include hemorrhages, neovascularization, bright lesions and laser treatment spots. Experts have labeled the blood vessels of the images. The dataset is mainly intended for the research of automated vessel segmentation but it is also valid for classification. The images were acquired with a Canon CF-60UVi camera and the image size is 3504*2336 pixels. The coverage of the retina is 60 degrees. (Odstrcilik et al. 2013)

3. MEDICAL IMAGE ANALYSIS

The human eye was described shortly in Chapter 2.2. First, we receive an image through our eyes, then it is transmitted through the nervous system and finally, it is processed and analyzed by our brain. A computer vision system acquires its images by a camera. Images can be defined as of a two-dimensional function *f(x,y)* with spatial coordinates *x* and *y*. The value of *f* is the gray level or color at a certain point *(x,y)* of the image. These points are called pixels and an image can be stored as an array of values *I[x,y]*. Algorithms that extract quantitative information are used to process and analyze images. The aim of the system is either to make a decision about the content of the image or to classify containing objects. (Shapiro & Stockman 2001; Nixon & Aguado 2012)

Generally, medical image analysis has four steps: preprocessing, region extracting, feature engineering and classification. The purpose of the preprocessing steps is to reduce noise, clear artefacts, enhance image quality and normalize images. Normalization is an important step for example when illumination conditions vary with images. Analyzing medical images also includes detecting different regions. Those regions can be extracted for better visualization and careful analysis. To better analyze the regions, quantitative measurements of the interest regions can be conducted. As a result, feature vectors are created. Finally, these regions can be analyzed and classified accordingly. In supervised learning, the outputs are known for the training samples and a classifier can be trained guided by that knowledge of outputs with the feature vectors as inputs. In medical image analysis, the known outputs are the ground truths made by medical experts. The results of the classifier are compared with the ground truths of the validation samples to determine the performance of the classifier. The topics are presented in the consideration of this study and its methods. (Shapiro & Stockman 2001)

3.1 Image preprocessing

Image preprocessing is an important step in image analysis. In digital images, pixels are finite and discrete values and images can be presented with standardized color models (Gonzalez & Woods 2002). To make the images more uniform color wise, image enhancement and normalization can be applied. Some image analysis applications need segmentation of objects or areas and this can be done with the help of binary operations.

3.1.1 Color models

A color model or color space describes and specifies colors in a standardized and commonly accepted way. Colors are bind to a coordinate system where a single point represents a certain color. Common color models in image processing are RGB, CMY and HSV models. (Gonzalez & Woods 2002)

A well-known color model is the RGB model. RGB stands for red, green and blue. This model is hardware oriented and it is most commonly used in color monitors. When an image is presented in the RGB color space, it is a combination of three component images. In Figure 6a there is a RGB image and Figures 6b, 6c and 6d represent the R, G and B component images. In the component images, the pixel value 0 represents black and pixel value 255 represents white when the pixel depth is 8 bits per channel and in total 24 bits. A 24-bit RGB image can have 16 777 216 colors but in the end only a few hundred colors are used in everyday applications. (Gonzalez & Woods 2002)

Figure 6. (a) RGB image and (b–d) three color channels

The CMY and the CMYK color models are hardware oriented and they are used in color printing. The letters stand for cyan, magenta, yellow and black. These colors (excluding black) are the primary colors of the pigments and secondary colors of light. The CMY model also uses three component images. The relationship between RGB and CMY is expressed as

$$\begin{bmatrix} C \\ M \\ Y \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} - \begin{bmatrix} R/255 \\ G/255 \\ B/255 \end{bmatrix}, \tag{1}$$

where R, G and B are first normalized to scale between zero and one. Resulting C, M and Y are also in the scale between zero and one. Black is added as fourth color in the CMYK model since it is a predominant color in printing. In theory, mixing equal amounts of cyan, magenta and yellow produces black, but for the printing purposes, this method is not functional. (Gonzalez & Woods 2002)

Another color model is the HSV model, which has been developed from HSI model that resembles the way that humans interpret and describe colors. The HSV stands for hue, saturation and value. The hue describes a pure color, the saturation corresponds to the

dissolved white light and the value describes brightness or intensity. The relationship between RGB and HSV is

$$H = \begin{cases} \theta, & B \leq G \\ 360 - \theta, & B > G \end{cases}, where \tag{2}$$

$$\theta = \cos^{-1}\left\{\frac{\frac{1}{2}[(R-G)+(R-B)]}{[(R-G)^2+(R-B)(G-B)]^{1/2}}\right\},$$

$$S = 1 - \frac{\min(R,G,B)}{\max(R,G,B)}, \tag{3}$$

$$V = \max(R,G,B)\,, \tag{4}$$

where H and θ are given as an angle that rotates counterclockwise. The HSV color space is advantageous for grayscale techniques since it separates grayscale and color information in an image. (Gonzalez & Woods 2002)

3.1.2 Contrast enhancement and color normalization

Contrast is the difference in luminance or color. If an image has low dynamic range of gray levels, the contrast is low. A simple method in order to get higher contrast is contrast stretching where the range of gray levels, between black and white, is broadened. Histogram represents the distribution of gray level values in an image. When the values are spread through the histogram, the result is fuller gray scale range. This method is called histogram equalization. (Gonzalez & Woods 2002)

Contrast-limited adaptive histogram equalization (CLAHE) is an advanced local contrast enhancement method. The method processes small tiles of an image and combines the results with bilinear interpolation to avoid local boundaries. In the beginning, a clip limit is set for the histogram of a gray level image depending on histogram normalization and neighborhood size. The output value for a single pixel is its rank in the local histogram. A part that exceeds the clip limit is redistributed to other bins and the clip limit is modified. The procedure can be repeated recursively if the results are not satisfactory. (Pizer et al. 1990)

When images that are compared come from multiple subjects, normalization may be needed for successful comparison. Here, normalization refers to the mapping of the colors of different images into a common norm. Varying illumination creates the distribution of color values in an image. Color normalization can be performed with image enhancement techniques such as histogram equalization, histogram specification or color channel

length normalization. In the histogram specification, the histograms are matched to a reference histogram. In color channel length normalization, the pixels are normalized first and then each color channel. (Goatman et al. 2003; Toennies 2012)

3.1.3 Segmentation and binary operations

In image segmentation, an image is subdivided into regions or objects for example into a background and an object in the image. Algorithms in image segmentation are generally divided into two basic styles: algorithms that make the partition based on sudden changes in intensity (edges) or based on regions that share certain criteria. A fundamental approach in both categories is the thresholding method. (Gonzalez & Woods 2002)

In its simplest form, global thresholding creates a binary image from an original grayscale image by a mathematical operation with a set threshold level. Thresholding turns all pixels that are below the given threshold level to zero (representing background points) and other pixels to one. The binary image can be used as a mask when segmenting areas or objects from the original image. The pixels of the original image are multiplied by the corresponding binary value of the binary mask. The pixels that were classified as object points stay unchanged while the pixels classified as background points are set to zero. (Gonzalez & Woods 2002)

Thresholding is effective when an image has a high level of contrast. The problem with the method is that it only considers intensities of pixels but not their relationships. It can be also sensitive to uneven illumination or noise. The threshold level can also change over an image and it could be adapted for example with the help of a histogram. To solve the uneven illumination, an adaptive thresholding method can be used using local properties. (Gonzalez & Woods 2002)

If the binary mask was created by thresholding, it may need some preprocessing for example with morphological methods before applying it to the original image. Two basic operations of morphology are erosion and dilation, where a structuring element is used to either enlarge or scale down a region (Figure 7). The structuring element is a neighborhood of binary pixels and it represents a shape in any size and chosen structure. Common shapes are a rectangle and circular shape, but also for example square, diamond, disk or octagon shapes could be used. (Shapiro & Stockman 2001; Gonzalez & Woods 2002)

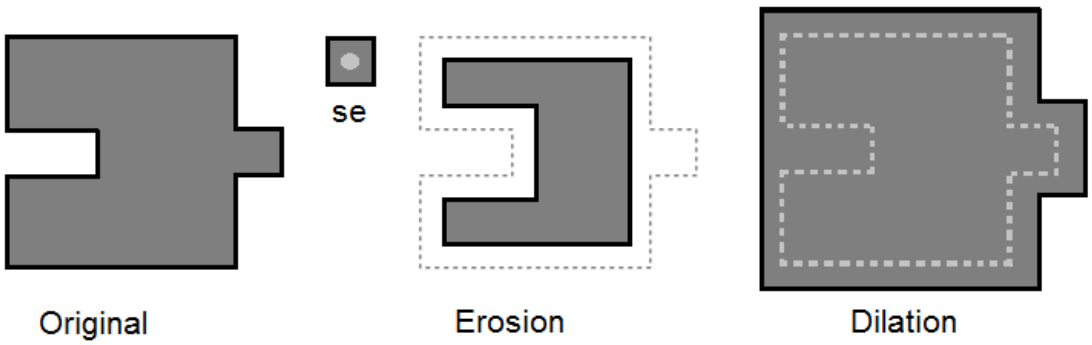

Figure 7. Morphological operations with the structuring element (se)

In Figure 7, the original object is processed with a square structuring element "se". The center point of the structuring element is traveling through the edges of the original object. In the erosion operation, the structuring element has traveled through the dotted line, studied the local neighborhood and erased object pixels where needed. The dilation operation follows same protocol but instead of erasing object pixels, the structuring element adds pixels. (Shapiro & Stockman 2001; Gonzalez & Woods 2002)

3.2 Feature extraction and texture features

Feature extraction is used to extract relevant information from images. To quantify the information, the properties of image texture can be used. A texture feature is in a form of feature vector, which includes a set of measurements from images.

Structure or a surface of an object has properties described as textures. For humans, the concept of texture seems obvious but it is quite challenging to define it. Whereas in digital images, a group of pixels in different tones, scales and structures represent different textures. One pixel alone cannot define texture. Image textures have information about amount and type of texture areas (primitives) and their spatial relationships. (Tuceryan & Jain 1998)

In computer vision, recognition and classification of textures are useful and important part of a process. Texture has been used for example in recognition of objects, shape analysis and image segmentation. A general challenge in computer vision is the shape of a target. Texture can give more information about the shape and surface orientation of an object. Texture segmentation tries to identify texture boundaries even though the actual textures are unknown. (Tuceryan & Jain 1998)

Textures can be described for example as coarse or fine. When texture elements are large and include several pixels, the texture is coarse. The texture is fine when the elements are small and neighboring pixels clearly differ in tone. (Sonka et al. 2008) Other possible properties to describe textures are a direction, directionality, frequency, a phase, density,

roughness, coarseness, regularity, uniformity and linearity. Some of the properties are dependent on each other. (Tuceryan & Jain 1998)

A geometrical method in texture analysis is based on the idea of texture primitives and structural approach assumes that texture has repeated or regular relationships between tones or structures. Textures do not always have geometrical regularity and cannot adequately be described by shapes. With statistical texture description methods, different statistical properties are computed from an image texture based on the gray level spatial distribution. It is suitable for images where pixel size and a texture primitive are comparable. Model based methods try to interpret a texture by generative image models and are more common in synthetic texture generation. (Tuceryan & Jain 1998; Sonka et al. 2008)

A texture feature is in a form of feature vector, which includes a set of measurements. A good feature vector assigns the texture to some specific class. First-order histogram features, gray level co-occurrence matrix, local binary pattern, Gabor filter and Laws' texture energy measures are all methods to extract texture features from images.

3.2.1 First-order histogram features

Variance, skewness and kurtosis are first-order features. These features are calculated using intensity histograms. When a histogram is used, variance measures the dispersion of the region intensity, skewness estimates the degree of asymmetry and kurtosis describes peaks in the histogram. (Batchelor & Whelan 2012) The equations are

$$Variance: \sigma^2 = \sum_{i=0}^{A-1} (i-\mu)^2 \, h(i) \,, \tag{5}$$

$$Skewness: \mu_3 = \sigma^{-3} \sum_{i=0}^{A-1} (i-\mu)^3 \, h(i) \,, \tag{6}$$

$$Kurtosis: \mu_4 = \sigma^{-4} \sum_{i=0}^{A-1} (i-\mu)^4 \, h(i) \,, \tag{7}$$

where i is a gray level value, A is the amount of gray levels in an image, h is the histogram, σ is standard deviation and μ is mean. These features can be used to describe texture but they lack information about the spatial relationships of pixels. (Batchelor & Whelan 2012)

3.2.2 Gray Level Co-occurrence Matrix (GLCM)

Gray level co-occurrence matrix (GLCM) is a statistical texture description method. It is based on a repeated occurrence of gray level structures in an image. The image can be quantized into a number of gray levels to limit the size of the resulting matrix. Co-occurrence matrix *CM* includes frequencies of two pixels' occurrence. It indicates how many times the reference gray level value *i* occurs with the neighbor gray level value *j* in the certain spatial relationship in an image *I* (Figure 8). The GLCM can be computed in the following way:

$$\boldsymbol{CM}(i,j) = \sum_{p=1}^{D}\sum_{q=1}^{E}\begin{cases} 1, & \text{if } I(p,q) = i \text{ and } I(p+\Delta x, q+\Delta y) = j \\ 0, & \text{otherwise} \end{cases} \quad (8)$$

where *D* and *E* indicate the size of an image *I*, Δx is displacement in rows and Δy is displacement in columns. The displacement can be also presented as an angle and a distance where horizontal is 0 degrees. The size of the matrix *CM* is the number of gray values to the power of two. Processing an image takes the number of operations that are proportional to an image size (to *D* and *E*). Size of the matrix can be reduced by counting mirrored gray level pairs together for example by adding up pair I(1,2) and I(2,1). (Haralick & Shanmugam 1973; Shapiro & Stockman 2001; Batchelor & Whelan 2012)

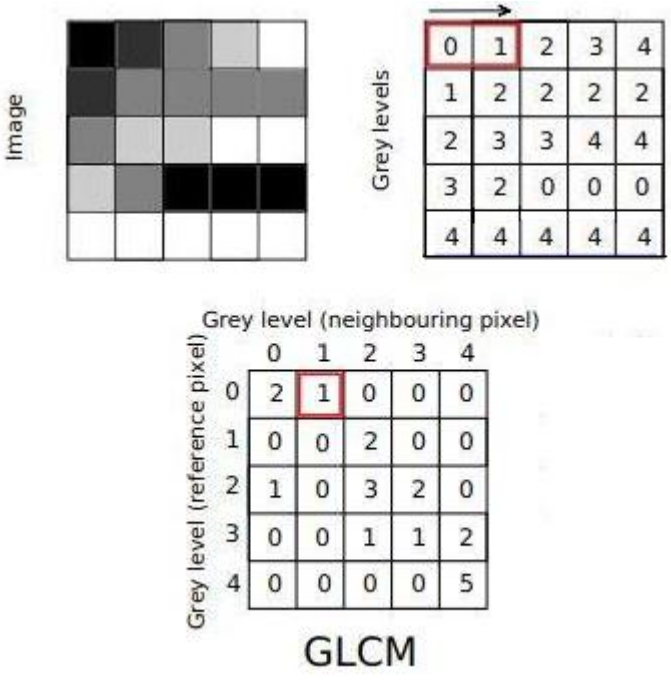

Figure 8. Creation of GLCM in horizontal direction (0°). Adapted from (Sikiö 2016).

Rapidly changing co-occurrence distribution within a distance indicates a fine texture. While GLCM gathers properties of this kind from a texture, it is not that useful in the resulting form for comparing textures. Several, more compact statistical parameters can

be calculated from the co-occurrence matrix such as energy, contrast, homogeneity, entropy, autocorrelation, dissimilarity, cluster shade, maximum probability, sum average, sum entropy, sum variance, difference variance, difference entropy and two information measure correlations. These parameters are presented in the Appendix A. Some of the Haralick texture parameters describe the nature and complexity of varying gray tone structures when others are related to more specific textural characteristics and organized structures. (Haralick & Shanmugam 1973; Shapiro & Stockman 2001) For example, energy measures image uniformity, entropy measures the complexity and contrast measures the local variation of intensity (Batchelor & Whelan 2012).

3.2.3 Local Binary Patterns (LBP)

Local binary patterns (LBP) method combines structural and statistical methods. LPB is very simple and it is based on pixel thresholding. Each pixel is compared with its neighboring pixels if their intensity is greater. In the beginning, the center pixel was compared with the neighborhood of size 3x3 but LBP was further developed to use circular neighborhoods in different sizes (Figure 9). The use of bilinear interpolation allows any number of pixels and any radius to be set. The method stores the occurrences of different patterns in the particular neighborhood of each pixel in to a histogram. The image can be divided into cells that all have their own histogram. The histograms of the cells are then combined to form the feature vector. LBP has invariance in rotation and illumination. It is computationally simple and therefore fast. (Ojala et al. 2002)

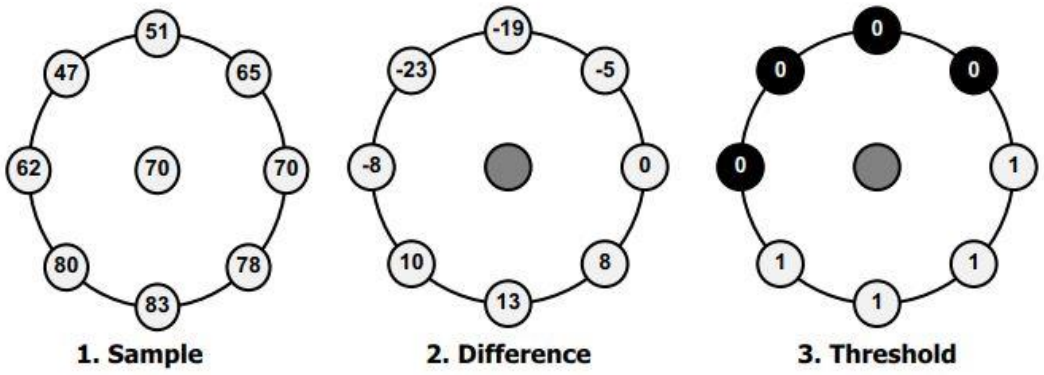

Figure 9. Creation of LBP. 1) The sample, 2) the calculated differences related to the center and 3) the threshold values. (Pietikäinen & Zhao 2009)

3.2.4 Gabor filters

The Gabor filter is a statistical method and originally it is a signal processing method. The Gaussian kernel function called the envelope is modulated by a plane wave that is complex and sinusoidal. The plane wave is called a carrier. A 2D Gabor filter *GF* in the spatial domain is defined as:

$$\boldsymbol{GF}(x,y) = \frac{f^2}{\pi\gamma\eta} * \ exp\left(-(\frac{f^2}{\gamma^2}x'^2 + \frac{f^2}{\eta^2}y'^2)\right) exp(j2\pi f x')\,, \tag{9}$$

$$x' = x\cos\theta + y\sin\theta\,, \tag{10}$$

$$y' = -\mathrm{x}\sin\theta + y\cos\theta\,, \tag{11}$$

where θ is the orientation of the surface normal to the parallel stripes, γ is the spatial aspect ratio (major), η is the spatial aspect ratio (minor) and f is the frequency of the sinusoid. The Gabor filters are convoluted with the original image. A Gabor filter is selective in frequency and orientation and it emphasizes edges and changing textures. For example, selecting eight orientations and five scales create 40 Gabor filters. The Gabor features are extracted from gray level images and further calculations can be conducted to get a compact presentation. (Petkov 1995; Haghighat et al. 2015) The method is robust against noise and illumination changes and it is invariant to translation, scale and rotation (Kamarainen et al. 2006).

3.2.5 Laws' texture energy measures

Laws' texture energy measures are based on local masks. Fixed-size windows are used for measuring the amount of variation within the window. Laws' texture energy masks are named as Wave, Level, Ripple, Edge and Spot. (Laws 1979; Sonka et al. 2008) Laws (1979) defined the center-weighted vector masks as:

$$\begin{aligned}
L_5\,(Level) &= [\ 1 \quad 4 \quad 6 \quad 4 \quad 1] \\
E_5\,(Edge) &= [-1 \quad -2 \quad 0 \quad 2 \quad 1] \\
S_5\,(Spot) &= [-1 \quad 0 \quad 2 \quad 0 \quad -1] \\
R_5\,(Ripple) &= [\ 1 \quad -4 \quad 6 \quad -4 \quad 1] \\
W_5\,(Wave) &= [-1 \quad 2 \quad 0 \quad -2 \quad 1]
\end{aligned}$$

The vectors are zero-sum (except Level), independent but not orthogonal. A set of 2D convolutional masks presented in Table 6 can be produced by taking the outer product of the 1D vectors. The product of *E5* and *L5* creates the mask *E5L5*. The Wave and its products are not commonly used. To make the method invariant to rotation, some of the symmetrical masks can be replaced with their average such as *E5L5* and *L5E5*. The method can be made invariant to changes in contrast if the convoluted images are normalized with *L5L5* mask. With filtering, the method can also be made invariant to luminance. (Laws 1979; Shapiro & Stockman 2001)

Table 6. Laws' energy maps

L5E5/E5L5	L5S5/S5L5	R5R5	R5W5/W5R5	E5W5/W5E5
L5R5/R5L5	E5E5	S5R5/R5S5	L5W5/W5L5	W5S5/S5W5
E5S5/S5E5	E5R5/R5E5	S5S5	W5W5	

The basic algorithm of Laws is based on local energy. First, an image is filtered with a set of convolution masks to produce a set of filtered images. Then, the texture energy is measured for each pixel by processing the image with the local texture energy filter, which is a moving window average of the absolute valued pixels. Finally, statistics can be computed from the filtered image by for example calculating sum of squared pixel values that are normalized by the number of pixels. (Laws 1979)

3.3 Feature selection

Feature selection is used to find the informative and optimal combinations of features. When hundreds of features are used in classification, this increases dimensionality and makes the classification computationally intensive. Therefore, it is more efficient to find features that have the best class separability and cut out features that are less efficient. This can lead to data reduction, increased processing speed and improved performance of classification.

One of the categories in feature selection is filter methods, which rank individual features using a relevance index based on statistics or correlation coefficients. These methods are simple and practical, but ignore the classifier interaction. On the other hand, features can be more useful in a subgroup even though their individual properties are not that decisive. Wrapper methods study feature subsets and use a learning machine to score the subsets according to their predictive capabilities. Wrappers can be computationally intensive (depending on the learning machine) and have high risk of overfitting. Embedded methods are more specific to a selected learning machine and feature selection is performed under a learning process. Hybrid methods combine above mentioned methods. (Guyon & Elisseeff 2006)

3.4 Classification and evaluation

Classification is used to assign class labels for the samples. The class labels could be for example “normal” and “abnormal”. Classification is done based on the features and a sample can have a large number of features. When there is large amount of features, the dimensionality of feature space is also large. In supervised learning, the labels for the samples are already known. The classification has two steps: training and testing. Purpose

of the training step is to tune the classifier parameters based on the input and output connection. After the training step, the testing step assesses the performance of the classifier. The classifier can predict classes for new data. It is important that the samples in the training set and in the testing set belong only to one set. Otherwise the results are not proper. (Haykin 1999; Shapiro & Stockman 2001)

The classification is done with a classifier. Classifiers are algorithms based on mathematical representations. There are several classifiers for two class problems and one of them is support vector machine (SVM). SVMs are feedforward networks that have hyperplanes as decision surfaces in a high dimensional feature space (Figure 10). When data is mapped into a higher dimensional space, it might be easier to separate the classes. SVMs aim to maximize the margin of separation between negative and positive samples. Support vectors are samples that the classifier has selected in the maximizing process and that are closest to the decision boundary. When the dataset is not separable, the aim is to minimize the probability of classification error. Adding a slack variable into the equation allows some of the point to be on the wrong side of the hyperplane. When the classes are not linearly separable, SVM can use kernels to create nonlinear decision surfaces. Such kernels are polynomial and radial-basis function (RBF). (Haykin 1999)

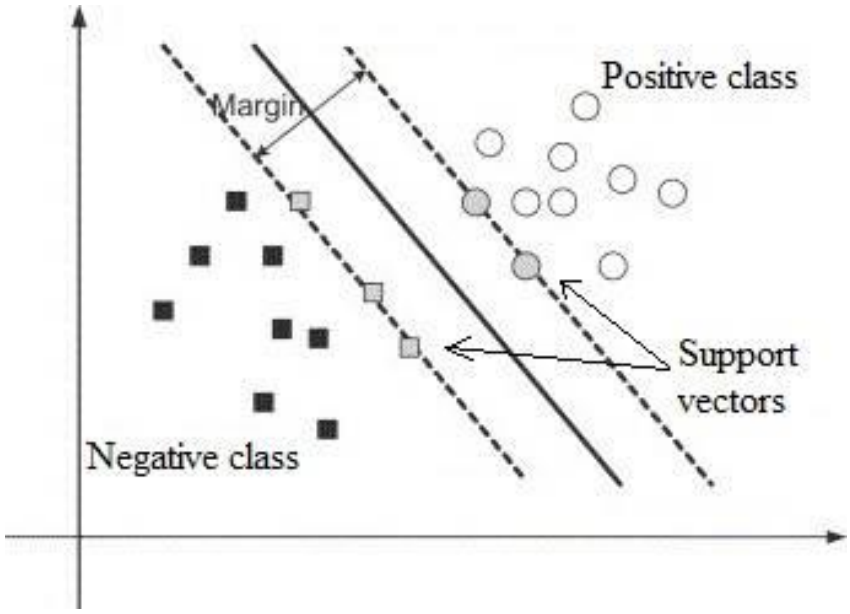

Figure 10. Support vector machine and the support vectors. Adapted from (Widodo & Yang 2007).

Another classifier is the k-nearest neighbors (kNN) algorithm. The classification is based on the distance of the nearest samples in the dimensional space. The assigned class for the new query instance is the majority vote of the classes of k-nearest instances. For example, in the Figure 11 the 5-nearest neighbor algorithm assigns the class for the new instance x_q as negative based on the five nearest instances. (Mitchell 1997)

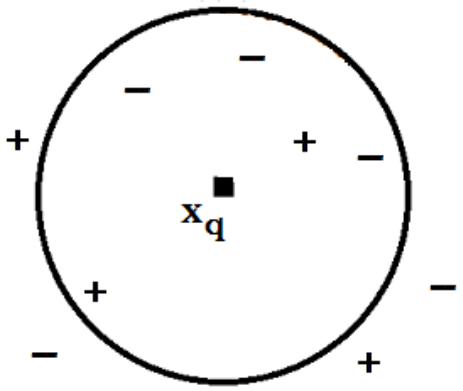

Figure 11. The query instance Xq would be assigned as negative with the 5-nearest neighbor (Mitchell 1997)

The classifier outputs are evaluated related to the ground truths. A ground truth for a sample is the correct class that has been assigned for example by a medical expert. If the ground truth and the test result both consider the sample as "abnormal", this is considered as true positive (TP) while mutual consideration for "normal" is known as true negative (TN) (Figure 12). If the ground truth is "normal" for a sample but the classifier output is "abnormal", it is considered as false positive (FP) and the opposite situation is false negative (FN).

		Ground truth	
		Abnormal	**Normal**
Test result	**Abnormal**	True positive (TP)	False positive (FP)
	Normal	False negative (FN)	True negative (TN)

Figure 12. Relationships between ground truths and test results

With the outcomes of TP, TN, FP and FN, the classification accuracy can be measured along with the sensitivity and specificity as:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}, \tag{12}$$

$$Sensitivity = \frac{TP}{TP + FN}, \tag{13}$$

$$Specificity = \frac{TN}{TN + FP}. \tag{14}$$

Accuracy, sensitivity and specificity are presented in percentages. The sensitivity presents how well the abnormal cases were classified correctly while the specificity presents this for the normal cases. If the specificity is high, it means that high percentage of the normal subjects are classified correctly as normal and only some samples are false alarms. The

receiver operating characteristic curve (ROC curve) method is used for classification method comparison. The curve plots the true positive rate against the false positive rate and the value of area under the curve (AUC) can be calculated from it.

The aim of the classifier is to classify the training samples accurately. A challenge arises when the training set is too small. This can lead to a tuned model that does not represent the data well enough. In overfitting, while the model is accurate on the training set, it performs worse with the test set. This can be the case also with noisy data. The ideal situation is that the classifier is designed so that it generalizes well for new data and it is accurate. (Mitchell 1997)

3.5 Analysis of retinal images

In this section, three methods for the retinal image analysis in the literature are presented. The literature is selected from the past five years based on their recognition. The methods are either used for red lesion detection or retinal image classification. Some methods are used for both purposes. Table 7 presents the summary of the results of the methods.

Roychowdhury et al. (2014) created a computer-aided screening system for diabetic retinopathy, which was able to detect red and bright lesions from fundus images and generate a severity grade for NPDR. Their system was based on three stages: image segmentation, lesion classification and severity grading. Before the first stage, the images were preprocessed with histogram equalization, contrast enhancement and filtering. This was done to eliminate artifacts, blurriness and uneven illumination. In the first stage, red lesion candidates and bright lesion candidates were detected as two foregrounds while the optic disc and vasculature were detected as a background. The vasculature was detected using shade-correction and region growing method whereas the optic disc was detected using pixel intensity and location. The background objects were extracted since an optic disc can be mistaken as a bright lesion and vasculature as red lesions. The second stage classified the red lesion and bright lesion candidates as non-lesions or true lesions. Then, the bright lesions were further classified as hard exudates or cotton wool spots and the red lesions were classified as microaneurysms or hemorrhages. They used 78 features including 14 structure-based features (area, orientation, solidity etc.), 16 features calculated from RGB and HSI color planes (minimum, maximum, mean and standard deviation) and 48 features using second-order derivative images. The structure-based features are primarily useful for classification between lesion and non-lesion. The features were scaled between 0–1. With feature selection and ranking process, they selected 30 features for classification. The used classifiers were Gaussian Mixture model for bright lesion detection and kNN for red lesion detection. The third stage counted the lesions and assigned a diabetic retinopathy severity grade for an image. The system classified red lesions and non-lesions in DIARETDB1 set with sensitivity 80 % and specificity 85 %. The results for bright lesions and non-lesions were slightly better with sensitivity 89 % and specificity 85 %. The method was also tested on Messidor dataset and trained with DIARETDB1

set. The best sensitivity was 100 %, specificity 53 % and 0.90 AUC with the test set of 1 200 images. (Roychowdhury et al. 2014)

Acharya et al. (2012) created a system that was able to classify fundus images into four classes: normal, NPDR, PDR or macular edema. They used Messidor dataset, but selected only 180 images in total. First, the images were preprocessed with image cropping and adaptive histogram equalization to remove the non-uniform illumination and then, the images were converted into grayscale images. They extracted several GLCM and run-length matrix based texture features from the images. However, they used only homogeneity and correlation from GLCM and short run emphasis, long run emphasis and run percentage from run-length matrix. The selection was based on analysis of variance (ANOVA) of the feature significance. The classification was done with SVM using either linear, RBF or polynomial kernels. With one-against-all method, they were able to use the SVM for multiclass classification. Testing used 3-fold validation method and the averages were calculated. The best average accuracy was 85 %, average sensitivity 99 %, average specificity 90 % and 0.97 AUC with the test set of 54 images. The best result was achieved with the 3-order polynomial kernel. (Acharya et al. 2012)

Antal & Hajdu (2012) created an ensemble-based framework for microaneurysm detection. The aim of the ensemble creation was to find the best combination of a preprocessing method and a candidate extractor. The preprocessing methods were Walter-Klein method, CLAHE, vessel removal and illumination equalization. They also considered the situation where there is no preprocessing done as “no preprocessing”. Walter-Klein method is used for contrast enhancement and the illumination equalization was implemented with vignette correction. The preprocessing was performed before the microaneurysm candidate extraction. The aim of the candidate extraction is to spot all objects with microaneurysm-like characteristics. The candidate extractor algorithms were called Walter, Spencer, Hough, Zhang and Lazar based on the authors of these methods. The algorithms were based on diameter closing, top-hat transformation, circular Hough-transformation, Gaussian mask matching and cross-section profile analysis. The ensemble creation had 25 pairs of preprocessing methods and candidate extractors. The number of pairs leads to 2^{25} possible combinations. They treated it as an optimization problem and used simulated annealing as the search algorithm. The ensemble system was used for microaneurysm detection and for diabetic retinopathy grading. The microaneurysm detection was evaluated with ROC database, DIARETDB and one private database. For the ROC database, combinations of Walter, Lazar and Zhang methods with all the preprocessing methods were selected for the ensemble. For the DIARETDB set, the selected methods were Walter, Lazar and Zhang with CLAHE, vessel removal and “no preprocessing”. The results were reported as free-response ROC curve (FROC curve) which corresponds to the located lesions. The system was also used for diabetic retinopathy grading and it was evaluated with Messidor database (that was trained with ROC database). The best accuracy was 82 %, sensitivity 76 %, specificity 88 % and 0.875 AUC with the test set of 1 200 images.

The paper lacks detailed information about the selected ensembles for Messidor set. (Antal & Hajdu 2012)

Table 7. Summary of the three recent classification results in the literature

Method for the red lesions detection	**Set**	**Accuracy**	**Sensitivity**	**Specificity**	**AUC**
Roychowdhury et al. (2014)	DIARETDB1	-	80 %	85 %	0.84
Antal & Hajdu (2012)	ROC	-	-	-	0.55
Method for the image grading	**Set**	**Accuracy**	**Sensitivity**	**Specificity**	**AUC**
Roychowdhury et al. (2014)	Messidor	-	100 %	53 %	0.90
Acharya et al (2012)	Messidor (180 images)	85 %	99 %	90 %	0.97
Antal & Hajdu (2012)	Messidor	82 %	76 %	88 %	0.88

4. METHODS

The aim of the system was to classify retinal images into normal or abnormal classes. In this case, abnormal can correspond either to diabetic retinopathy in general or microaneurysms that are a symptom of diabetic retinopathy. To classify the images into two classes, features were used. Features are numerical descriptions and input variables for the classifier. Texture features are based on image pixel information and they are used to find and describe patterns or elements. After image preprocessing, texture features were extracted from the images and used for classification. The work was conducted with Matlab® R2016a software.

The presented system has advantages compared with the methods presented in the literature. Roychowdhury et al (2014) have a complicated system with several steps and their system extracts vasculature, an optic disc, red lesion and bright lesion candidates. The number of steps increases the complexity and possibility of errors. Our system is simple and has adequate computational cost. While Acharya et al. (2012) had only a small testing set, our system is tested with a bigger test set and with the 5-fold validation method. In addition, our system is able to process images that have different sizes and FOVs.

4.1 General processing pipeline

Figure 13 describes the processing pipeline. First, images were preprocessed with resizing, background segmentation and contrast enhancement. Second, texture features were extracted from the green color channel of every RGB image. Then, the features were normalized, ranked, and fed into a feature selector to find and select the ones that had the biggest difference between the two classes. Finally, the selected features were fed into a SVM classifier. The SVM outputs for the samples were class 0 as healthy or class 1 as diabetic retinopathy. As presented in Table 4, the class 0 samples do not have microaneurysm or hemorrhage findings as opposed to the class 1.

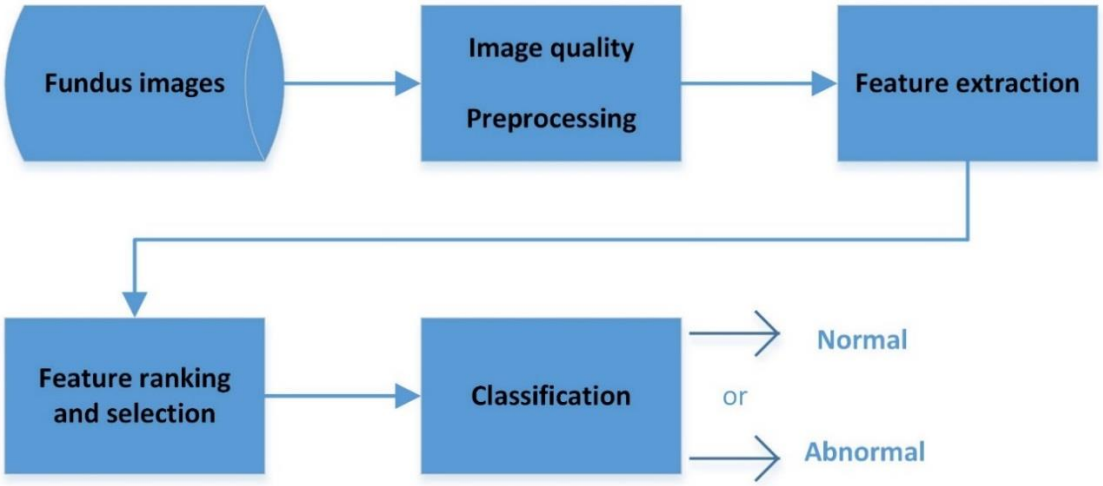

Figure 13. Processing pipeline of the system

Another experiment used image blocks instead of whole images. When the images were divided into blocks, the process had more steps. First, images were preprocessed with resizing, background segmentation, color equalization, contrast enhancement and optic disc removal. Second, the images were divided into blocks. If a block included background, then the block was discarded. Third, the features were extracted from the green channel of the RGB image blocks. Then, the features were normalized, ranked and selected. Finally, the features were fed into a SVM classifier. The SVM outputs for the block samples were class 0 as healthy or class 1 as microaneurysm or small hemorrhage.

4.2 Image preprocessing

First step in image processing was image resizing and cropping. If an image dataset had images in different sizes, these images were resized to a uniform size. Retinal images tend to have lots of excess background and therefore the images were cropped near the retina to have less of it. Resizing images into a smaller size can lead to a loss of information but on the other hand, it will lower the computational workload. If the ground truth was presented as a binary image, that image was also resized and cropped to maintain the correlation to the retinal image.

A dark background was extracted with segmentation (Algorithm 1) to ensure that all images had the same values for it. This method would extract a part of a retina if some parts of the retina were close to the color of the background.

```
% Extract V-channel from HSV color space
hsv_all = rgb2hsv(original_image);
hsv_V = hsv_all(:,:,3);

% Create and fine tune a binary image
binary = hsv_V > 0.1;
binary = imfill(binary,'holes');
se = strel('square',5);
binary = imerode(binary,se);

% Create a 3-channel binary image
% and mask the original image
binim(:,:,1) = binary;
binim(:,:,2) = binary;
binim(:,:,3) = binary;
original_image(~binim) = 0;
```

Algorithm 1. Background segmentation with binary mask in Matlab

The Algorithm 1 has the following steps. First, a RGB image was converted into a HSV image and the value channel was saved as a variable *hsv_V*. Then, a binary image was created by thresholding *hsv_V* with 0.1 as the set limit. The assumed background was now marked as "0" and the assumed object as "1" in the binary image. Next, the binary image was processed by filling black holes inside the object area to make the area uniform. In addition, morphological erosion with the square structuring element *se* was used to erase small white particles and make the background uniform. Finally, the binary image (Figure 14b) was used as a mask to the original image with logical NOT-operation as "if a pixel is not an object point, it will be set to zero".

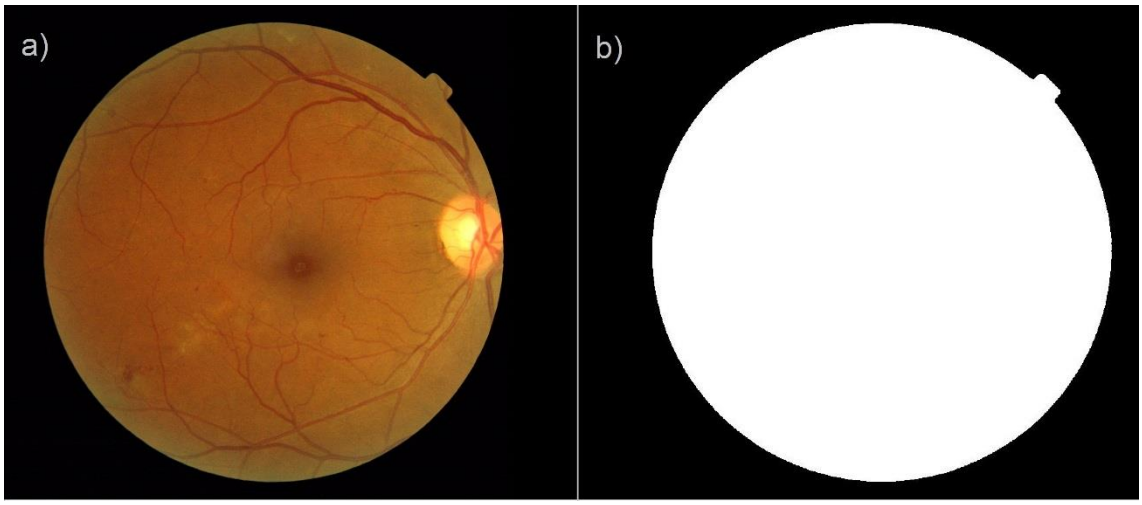

Figure 14. (a) The original retinal image and (b) its binary image

After the background segmentation, CLAHE processing was applied (Pizer et al. 1990). CLAHE was used to enhance the contrast in the images and it was also used by Hijazi et al. (2012), Acharaya et al. (2016) and Antal & Hajdu (2012). CLAHE was applied to the images with the Matlab function *adapthisteq*. The green channel of the RGB image was used in the following steps, because it has higher contrast regarding red lesions and retinal

blood vessels compared with the other channels in RGB. Use of the green channel is a common method in the literature (Reza & Eswaran 2011; Antal & Hajdu 2012; Roychowdhury et al. 2014; Acharya et al. 2016).

Experiments also used image blocks instead of whole images. Before the CLAHE method, the colors of the images were normalized to lower the variance between the images. First, a new image was created by dividing the pixels of the original grey level image with the averages of local windows. Then, the color channels of the original image are masked with the new image to correct uneven illumination. Finally, the color channels are tuned based on the top intensity values. Riku Hietaniemi implemented the algorithm for color normalization. After the color normalization, the optic disc was segmented from the images by hand. The optic disc was removed because it does not include microaneurysms and it can mislead the classifier. Then, the images were divided into blocks either by based on the lesion points or gridding the whole image with a set overlap of 50%.

4.3 Texture features

The texture features were used to extract information from the preprocessed images. The list of the texture features used in this work is presented in Table 8. The texture features were selected based on their availability.

Table 8. List of the texture features and their vector lengths

#	Feature name	Length
1	Variance	1
2	Skewness	1
3	Kurtosis	1
4	Energy	3
5	Contrast	3
6	Homogeneity	3
7	Entropy	3
8	Autocorrelation	3
9	Dissimilarity	3
10	Cluster shade	3
11	Maximum probability	3

#	Feature name	Length
12	Sum average	3
13	Sum entropy	3
14	Sum variance	3
15	Difference variance	3
16	Difference entropy	3
17	Information measure of correlation 1	3
18	Information measure of correlation 2	3
19	LBP	59
20	Gabor	48
21	Laws	14

The variance, skewness and kurtosis were calculated based on the Equations 5–7. The Matlab function *graycomatrix* was used to create the gray level co-occurrence matrix and the Matlab function *GLCM_Features4* from Uppuluri (2010) was used to calculate several statistical parameters from the created GLCM. The displacements were 45, 90 and 135 degrees with the distance of 1 pixel. The pairings were enabled which counts the mirrored gray level pairs together. To create the LBP feature vector, the Matlab function *extractLBPFeatures* was used. The radius was set to 4 and the neighborhood was set to 8. To create the Gabor filter feature vector, the Matlab script from Haghighat et al. (2015) was used. They did research on face recognition and they used the Matlab script for that purpose. They had set η, γ and maximum frequency values. We used four scales, six orientations, five rows and five columns in the Gabor feature bank. The selected values for the Gabor filter are presented in the Table 9. The feature vector included the means and standard deviations of the filtered images. Fourteen Laws' texture energy masks were used to produce the feature vector: L5E5, E5L5 L5S5, S5L5, L5R5, R5L5, E5S5, S5E5, E5R5, R5E5, S5S5, S5R5, R5S5 and R5R5. The masks including W5 and the mask L5L5

were left out. The L5L5 mask is usually used to normalize the images and the Wave filters are not commonly used.

Table 9. The selected values for the Gabor filter

$\gamma = \eta = \sqrt{2}$	$f_{max} = 0.25$	$f = \frac{f_{max}}{\sqrt{2}^{u}}$	$\theta = \frac{v}{6}\pi$	$u = 0 \ldots 3$	$v = 0 \ldots 5$

4.4 Feature selection and classification

The features had different scales and although they were comparable, the features were normalized to a scale from 0 to 1. The feature selection was implemented as a hybrid method. The normalized features were ranked by a class separability using the Matlab function *rankfeatures*. The function returns the ranked list of indices to the corresponding features. The final feature selection was done with a sequential feature selection algorithm, which belongs to wrapper methods. The method sequentially adds ranked features to the learning machine (i.e. SVM classifier) until certain stopping conditions are met. The algorithm stopped if no improvement in accuracy was observed. The method is based on forward search and it explores the combinations of features. The algorithm was computationally intensive due to the SVM classifier.

The selected classifier for the final image classification was a SVM with the RBF kernel. The SVM is an efficient classifier and RBF has good characteristics of creating nonlinear decision surfaces. It was used by Hijazi et al. (2012), Noronha et al. (2012), Krishnan & Faust (2013) and Acharya et al. (2012; 2016). The parameters of the kernel scale and box constraint (the cost of misclassification) were tuned to find the best fit.

5. EXPERIMENTS AND RESULTS

Three image datasets were selected for the testing and evaluation of the system. In this chapter, three experiments and the results are presented. The results are followed by the discussion. Chapter 4 described the used methods and settings in detail.

5.1 Selected datasets

Three datasets were selected for the experiments. Messidor dataset provides 1 200 images for the classification of the whole image into healthy or diabetic retinopathy. This dataset was used in the image classification in the global setting. DIARETDB1 and e-ophta datasets were used for the classification of the block images into microaneurysm or normal. These datasets provide coordinates for microaneurysms and small hemorrhages in the images.

ROC and DRiDB datasets were also suitable for the block image experiments but they have smaller amounts of images with ground truths than the selected image sets. Kaggle set has over 35 000 images with labels but the variations in quality, equipment and size within the images and the amount of noise are high. HRF dataset was too small for the whole image classification purposes.

5.2 Evaluation

The evaluation of the system is done on image level rather than pixel level. Therefore, a whole image or an image block has a ground truth and a classification result. The final results for the system are reported as average accuracy, sensitivity and specificity.

The 5-fold method was used to estimate the performance of the system. The K-fold method divides the sample set into subsets of $K-1$ training subsets and one testing subset. Then the testing is repeated K times with each subset used as testing set and the others as training sets. For example, the 5-fold would divide the sample set into four training subsets and one testing subset. The step is repeated five times and the average is calculated from the results. (Mitchell 1997, p. 112)

5.3 Results

There were three experiments. The first experiment studied the classification of the whole images into healthy or diabetic retinopathy in Messidor data set. The second experiment divided the images into blocks and studied the classification of the blocks into normal or microaneurysm in DIARETDB1 set and in e-ophta set. The third experiment studied the effects of the size of the image block in classification.

5.3.1 Experiment 1: Global image features

The first experiment studied the classification of whole images. The SVM with RBF kernel was used. There are 1 200 images in the Messidor set: 546 healthy (class 0) and 645 retinopathy images (retinopathy severity 1, 2 and 3 assigned to class 1). The size of the test sets was 240 images. There were in total 21 feature types and first, every feature type was tested independently. Table 10 presents the resulting average accuracy for the feature types with the 5-fold method. The best average accuracy was 65 % with LBP and with Laws. Average accuracy with Gabor feature was 63 %. Other features had average accuracies between 56 % and 60 %. The by chance accuracy would be 55 % if the classifier would assign all the images into the class 1.

Table 10. Results for the feature types in Messidor set

Feature name	Accuracy	Feature name	Accuracy
Variance	60 %	Sum average	58 %
Skewness	58 %	Sum entropy	57 %
Kurtosis	59 %	Sum variance	58 %
Energy	57 %	Difference variance	59 %
Contrast	59 %	Difference entropy	57 %
Homogeneity	56 %	Information measure of correlation 1	56 %
Entropy	56 %	Information measure of correlation 2	59 %
Autocorrelation	59 %	LBP	**65 %**
Dissimilarity	57 %	Gabor	63 %
Cluster shade	57 %	Laws	**65 %**
Maximum probability	58 %		

Then the features were combined based on the feature ranking and selection process. First, the feature selection was done with all the features that were ranked and either top ones or all the ranked features were fed into the feature selector. Next, the feature selection was done with only part of the features that were ranked and the top ones were selected for the feature selector. At this point, all the values in the feature vector were treated as independent features. Table 11 presents the resulting average accuracy, sensitivity and

specificity for different combinations with the 5-fold method. The best average accuracy of 69 % was achieved with the combination of 28 features including LBP, Laws and Gabor. Average accuracy of 68 % was achieved with 36 features including sum average, autocorrelation, difference entropy, LBP, Laws and Gabor. When the top 20 of all features were used in feature selection, the average accuracy was 65 % with 11 features including LBP and kurtosis. Ranking was also tested for only a part of features excluding Gabor and Laws. Selecting the top 20 and top 30 of these features gave average accuracy of 66 % with 20 and 11 features. The set of 20 features included LBP features and the set of 11 features included LBP, kurtosis and skewness. The size of the test sets was 240 images.

Table 11. Results for the selected features in Messidor set

Features fed into the feature selector	**Number of selected features**	**Accuracy**	**Sensitivity**	**Specificity**
All ranked, top 80	28	**69 %**	72 %	66 %
All ranked	36	68 %	70 %	65 %
All ranked, top 20	11	65 %	72 %	56 %
Ranked part, top 30	20	66 %	71 %	60 %
Ranked part, top 20	11	66 %	73 %	58 %

Messidor set has three classes for the stage of the diabetic retinopathy and earlier tests combined these classes into one class. The final test for the whole images was with separate severity classes. Classification was done with 28 features including LBP, Laws and Gabor. Table 12 presents the results for average accuracy, sensitivity and specificity between the severity classes. “Class 0” has 546 healthy images and “Class 3” has 254 retinopathy images in severity stage 3. With these two classes, the 5-fold method gave average accuracy of 82 % with a test set of 160 images. “Class 0” versus “Class 2” had average 5-fold accuracy of 71 % with 247 “class 2” images and a test set of 159 images. “Class 0” versus “Class 1” had average 5-fold accuracy of 80 % with 153 “class 1” images and a test set of 140 images. However, while the specificity was high (94 %, 94 % and 100 %), the sensitivity was low (59 %, 18 % and 6 %) for all the results.

Table 12. Results between the severity classes in Messidor set

Classes	Number of images	Accuracy	Sensitivity	Specificity
0 vs. 3	546 vs. 254	82 %	59 %	94 %
0 vs. 2	546 vs. 247	71 %	18 %	94 %
0 vs. 1	546 vs. 153	80 %	6 %	100 %

5.3.2 Experiment 2: Image blocks and normalization

Images were divided into blocks of size 150*150. The 5-fold tests were done 20 times to see the true effects of normalization and the results are averages of the measurements. The DIARETDB1 set had 397 normal block images and 463 microaneurysm block images. The microaneurysm blocks were generated based on the ground truth lesion points with overlapping blocks and the normal blocks were generated with overlapping gridding. The SVM classifier with RBF kernel was used.

The results with different feature sets for normalized and non-normalized images are presented in the Table 13. The feature set F1 included 28 features from LBP, Laws and Gabor and the average accuracy was 88 % for both normalized and non-normalized images. The set F2 included 36 features from LBP, Laws, sum average, autocorrelation, difference entropy and Gabor. The average accuracy with the set F2 was 90 % for normalized and 84 % for non-normalized images. The set F3 included 22 features from LBP and Laws. The average accuracy with the set F3 was 92 % for normalized and 88 % for non-normalized images. The set F4 included 30 features from sum average, autocorrelation, sum variance, Gabor, LBP and Laws. The average accuracy with the set F4 was 90 % for normalized and 84 % for non-normalized images. The set F5 included 30 features from LBP and Laws. The average accuracy with the set F5 was 91 % for normalized and 90 % for non-normalized images. Normalized images with the feature set F3 had the best accuracy, sensitivity and specificity (92 %, 92 % and 92 %). For the non-normalized images, the best accuracy was 90 % with sensitivity 89 % and specificity 91 % was with the feature set F5.

Table 13. Results for normalized and non-normalized image blocks in DIARETDB1

Normalization and feature set	Accuracy	Sensitivity	Specificity
Non-normalized Set F1	88 %	88 %	89 %
Normalized Set F1	88 %	89 %	88 %
Non-normalized Set F2	84 %	84 %	85 %
Normalized Set F2	90 %	90 %	91 %
Non-normalized Set F3	88 %	88 %	88 %
Normalized Set F3	**92 %**	**92 %**	**92 %**
Non-normalized Set F4	88 %	87 %	88 %
Normalized Set F4	90 %	90 %	90 %
Non-normalized Set F5	90 %	89 %	91 %
Normalized Set F5	91 %	92 %	90 %

The e-ophta set had 402 normal block images and 482 microaneurysm block images. The microaneurysm blocks and the normal blocks were generated by dividing the images with 50 % overlap. The ground truth for a block was based on the provided binary image.

The results with different feature sets for normalized and non-normalized images are presented in the Table 15. The feature sets F1 and F2 are the previously presented feature sets. The average accuracy with the set F1 was 82 % for normalized and 78 % for non-normalized images and the average accuracy with the set F2 was 86 % for normalized and 82 % for non-normalized images. The set F6 included 19 features from difference entropy, dissimilarity, contrast, homogeneity, difference variance, LBP and Laws. The average accuracy with the set F6 was 80 % for normalized and 76 % for non-normalized images. The set F7 included 30 features from LBP, sum variance, sum average, Gabor and Laws. The average accuracy with the set F7 was 86 % for normalized and 81 % for non-normalized images. The set F8 included 30 features from LBP, difference entropy, dissimilarity, homogeneity, contrast, difference variance, entropy and Laws. The average accuracy with the set F8 was 81 % for normalized and 78 % for non-normalized images. The best results were the product of the feature set F2. Normalized images had the best accuracy, sensitivity and specificity with 86 %, 87 % and 86 %. For the non-normalized images, the best accuracy was 82 % with 84 % sensitivity and 79 % specificity.

Table 14. Descriptions of the feature sets

Feature set	Size	Features
F1	28	LBP, Laws and Gabor
F2	36	LBP, Laws, sum average, autocorrelation, difference entropy and Gabor
F3	22	LBP and Laws
F4	30	Sum average, autocorrelation, sum variance, Gabor, LBP and Laws
F5	30	LBP and Laws
F6	19	Difference entropy, dissimilarity, contrast, homogeneity, difference variance, LBP and Laws
F7	30	LBP, sum variance, sum average, Gabor and Laws
F8	30	LBP, difference entropy, dissimilarity, homogeneity, contrast, difference variance, entropy and Laws

Table 15. Results for image blocks in e-ophta

Normalization and feature set	Accuracy	Sensitivity	Specificity
Non-normalized Set F1	78 %	82 %	73 %
Normalized Set F1	82 %	84 %	80 %
Non-normalized Set F2	82 %	84 %	79 %
Normalized Set F2	**86 %**	**87 %**	**86 %**
Non-normalized Set F6	76 %	80 %	70 %
Normalized Set F6	80 %	81 %	79 %
Non-normalized Set F7	81 %	83 %	78 %
Normalized Set F7	86 %	86 %	86 %
Non-normalized Set F8	78 %	80 %	75 %
Normalized Set F8	81 %	81 %	81 %

5.3.3 Experiment 3: Image block size

In experiment two, the images were divided into blocks of size 150*150. The experiment three studied the effects of the block size in e-ophta dataset. The images were normalized. The 5-fold validation method was used. The SVM classifier with RBF kernel was used.

Table 16 presents the results for the classification of different block size images. The block size 150*150 was now tested with a bigger set of images. There were 1 345 microaneurysm block images and 1 345 normal block images. Size of the test set was 538 images. The block size 100*100 had 1 680 microaneurysm block images and 1 680 normal block images with the test set of 672 images. The block size 75*75 had 1 949 microaneurysm block images and 1 949 normal block images with the test set of 780 images. The block size 50*50 had 2 119 microaneurysm block images and 2 119 normal block images with the test set of 848 images.

The sets F1, F2, F7 and F8 were used to test the effects of the block size. The sets were described in the previous subchapters. The best result was achieved with the smallest block size and with the feature set F2 where the average accuracy, sensitivity and specificity were 96 %. Also for the other block sizes, the feature set F2 gave the best results. The results of the other feature sets were uniform.

Table 16. Results for different block sizes in E-ophta

	Average accuracy / sensitivity / specificity	
Block size	**Set F1**	**Set F2**
150*150	82 % / 80 % / 85 %	86 % / 83 % / 90 %
100*100	87 % / 83 % / 91 %	90 % / 87 % / 93 %
75*75	87 % / 83 % / 91 %	92 % / 91 % / 93 %
50*50	93 % / 93 % / 93 %	**96 % / 96 % / 96 %**
Block size	**Set F7**	**Set F8**
150*150	82 % / 77 % / 86 %	81 % / 80 % / 83 %
100*100	86 % / 81 % / 91 %	86 % / 82 % / 90 %
75*75	87 % / 84 % / 89 %	87 % / 85 % / 89 %
50*50	93 % / 93 % / 94 %	93 % / 92 % / 95 %

5.3.4 Summary

The features were successful in local setting when the size of the image block was small. The best results came from e-ophta set with the block size of 50*50. This was due to the fact that the features were able to extract finer details (such as microaneurysms) on smaller images. On Messidor image set, the global setting was not successful. In addition, the variation inside the diabetic retinopathy class was significant and it affected the results. The color normalization method created slightly better results compared with the non-normalized images.

5.4 Discussion

The first experiment studied the classification of whole images into healthy or diabetic retinopathy using Messidor image set. The Messidor set is divided into healthy images and into three severity stages of diabetic retinopathy. The best accuracies for the whole image classification (healthy vs. retinopathy) were 65 % with the feature types tested independently and 69 % with the combinations of selected features. Sensitivity of the latter was 72 % and specificity 66 %.

When the features describe an entire image, the features are global. Global features compress the information on the images into a compact form. One reason for the low accuracy of the classification is that the information and smaller details get lost when a whole image is described in such compact form. Another challenge was that Messidor set has images in three different sizes and the images were resized into smaller, uniform size. Resizing an image into smaller size can lead to loss of information particularly on small details such as microaneurysms. The e-ophta set also has images in three different sizes while the DIARETDB1 is consistent in the size.

The fact that three severity stages were combined into one class affects the accuracy of the classification. Combining the classes creates lots of variation within the final class, which leads into decreased class separability. When the severity stages were compared individually with the healthy images, the accuracies were good but the sensitivities were low. One reason for this was the amount of healthy images compared with the amount of retinopathy images. For the severity class 3 and class 2, there were twice as much healthy images and compared with the class 1, the amount of healthy images was even bigger. Therefore, the healthy images were overpowering the classification and the amount of false negatives was considerable. This is specific to SVM classifiers. Roychowdhury et al. (2014) were able to get high sensitivity but with the cost of low specificity. Neither situation is favorable, but high sensitivity ensures that fewer people with the disease are missed. On the other hand, the amount of false negatives and the sensitivities that Roychowdhury et al. (2014) reported were not consistent. Therefore, their results may include some errors. Our results were also lower than what Antal & Hajdu (2012) and Achraya et al. (2012) have presented. On the other hand, Achraya et al. (2012) used only 180 images

from Messidor set and their test set was 54 images while our test set had 240 images. Such a small test set does not give a sufficient impression about their system.

The second experiment studied the classification of the image blocks into normal or microaneurysm using DIARETDB1 and e-ophta image sets. The size of the blocks was 150*150. The experiment also compared the effects of the color normalization method with the accuracy. For DIARETDB1, the best accuracy, sensitivity and specificity were 92 % with the normalized set of blocks. The used feature set F3 had 22 features including LBP and Laws. The best average accuracy, sensitivity and specificity for non-normalized images were 90 %, 89 % and 91 % with the feature set F5 that included 30 features of LBP and Laws. Both of the features are known to be strong and descriptive. For e-ophta set, the results were lower. The best results were measured with the feature set F2 that had 36 features and combined LBP, Laws and co-occurrence matrix related features. For the normalized images, the best accuracy was 86 % with 87 % sensitivity and 86 % specificity and for the non-normalized images, the best accuracy was 82 % with 84 % sensitivity and 79 % specificity.

The feature sets F1 and F2 were generated earlier with the feature selector during the first experiment, whereas the feature sets from F3 to F8 were generated solely for the image sets at hand. For the DIARETDB1 set, the combinations of LBP and Laws features were successful. These feature sets were generated for the image set. For e-ophta set, the feature selector did not produce a similar set of features and the generated feature sets F6, F7 and F8 were not as successful. Although the feature sets F6–F8 included LBP and Laws features, they also included weaker features. It is possible that the feature selection process has hit a local maximum and the process was misled. However, when using the image blocks the features are local rather than global. Now, the features can give more accurate information about the images.

Normalized images produced better results than non-normalized images. The difference in classification accuracy between normalized and non-normalized images was not great on average but in e-ophta set, the specificity was notable improved with normalization. The aim of the normalization was to correct uniform illumination and this have been advantageous especially for the classification of healthy images. Another reason may be that the DIARETDB1 set was acquired with one fundus camera at one place while the e-ophta set was collected from a set of 51 741 images. The e-ophta images were acquired with two different cameras at different places. Therefore, the DIARETDB1 set could be assumed to be quite uniform originally which is favorable for the classification.

The third experiment studied the effects of the block size on the classification. The testing was done with the e-ophta set and the sizes of 150*150, 100*100, 75*75 and 50*50 were tested with four sets of features. The best average accuracy was 96 % with 96 % sensitivity and 96 % specificity with block size 50*50 and with the feature set F2. The difference

from the results with the block size 150*150 was significant. When the size of the block image is small, the features can extract finer details such as small microaneurysms.

National Institute for Health and Clinical Excellence (NICE) of the UK and Diabetes UK have set guidelines for diabetic retinopathy screening. The sensitivity should be at least 80 % and the specificity should be at least 95 %. These are also known as Exeter Standards. (British Diabetic Association 1997) Our system was able to meet the requirements when considering the microaneurysm detection but the results on disease or no-disease related detection with our microaneurysm detector are unknown. The comparison of state-of-the-art methods is challenging due to different evaluation criterion. For example, Antal & Hajdu (2012) present their results for red lesion detection as partial AUC and their evaluation is pixel based rather than image based.

The e-ophta set block images were produced with the overlap of 50 %. To prevent the situation where microaneurysms were in between the gridded images and could have been missed, the overlap was used. This could have affected the classification results favorably since some images share the same areas. The use of a cross-validation method i.e. the 5-fold method gives more realistic results and a more accurate estimate for the system and its capabilities to generalize.

Image blocks that included parts of black background were removed from the classification training and testing set. This is a challenge since some parts of the retina may not be examined if certain image blocks are removed. When the block size is small enough, it increases the probability that all parts of the retina are examined. On the other hand, small block size increases the total amount of images and therefore it increases the computational burden. In addition, the number of normal images increased more rapidly than the amount of microaneurysm images when the size of the block was decreased. If the number of normal images overpowers the microaneurysm images in the training phase, this can lead to low classification sensitivity.

The system is limited to the microaneurysms and small hemorrhages. Fleming et al. (2010) state that the disease or no disease classification can be performed using only image quality assessment and microaneurysm detection. However, the automated detection of hemorrhages and exudates improved the detection of diabetic retinopathy in their study. Adding the detection, they managed to increase the sensitivity from 94.9% to 96.6%. Further research would be needed for the implementation of exudate and bigger hemorrhage detection to our system. Additionally, the system has not been tested with images that include laser treatment marks or other misleading markings. Therefore, it is unknown how our system behaves on such artefacts.

Visual detection and classification are challenging tasks since images tend to have great amount of variation. In retinal imaging, this variance comes from different cameras, the field of views, light, obstruction or the image quality is affected by blurring, focus, noise

or illumination. Our system has shown promising results both on uniform DIARETDB1 set and on non-uniform e-ophta set. The system excludes too dark parts of the retinal image and with preprocessing methods, fights the effects of non-uniform illumination.

Some of the textures were not that strong or descriptive. The first-order histogram features were too weak and the GLCM features had quite low success. Variance, skewness and kurtosis only describe the intensity of an image and lack the spatial information. The reason for GLCM features being weak might be that the Matlab code from Uppuluri (2016) had some challenges with efficiency and other features just performed better than the GLCM based features. Inexperience with the Gabor filters was a probable reason why the feature turned out to be not that descriptive for the task. On the other hand, it was included in the winning feature set, but it might function even better with proper settings. LBP features are computationally simple (Pietikäinen et al. 2011, p.4) and known to be descriptive. It did not come as surprise that LBP performed very well on our system. In addition, Laws' texture energy features performed well and were easy to implement. Advantage of simple features is a lower computational cost.

One limiting factor in this study was the lack of computational power. Training a SVM network has high computational workload and therefore the training took a lot of time. Especially the feature selection process which used the SVM classifier. The final system itself is not computationally intensive, since the network is already trained and the features are already selected. When a user gives an image as an input to the system, the image is first preprocessed with resizing, background segmentation, contrast enhancement and color equalization. Second, the user must select four optic disc boundaries for the optic disc extraction. Then, the image is divided into blocks and the blocks with the background are discarded. Finally, the selected features are extracted from the blocks and fed into the classifier. The classifier gives the result of healthy or microaneurysm for each image block.

The system is not fully automated since the user must select the boundaries of an optic disc. For future development, an optic disc detector could be implemented. This would save the time of the user and make the system more efficient. In addition, an image quality detector could be implemented to warn the user about images with unacceptable quality and inform the user about inadequate clarity or field definition as Fleming et al. (2006; 2010) suggested. The images used in this work are somewhat outdated. Messidor and DIARETDB1 image sets were acquired before 2007 and the e-ophta set in 2008–2009 (Kauppi et al. 2007; Decencière et al. 2014; MESSIDOR 2016). In ten years, fundus cameras have developed and offer better image resolution and increased sensitivity. The system should be tested with more recent retinal images.

6. CONCLUSION

The number of people with diabetes is constantly increasing and so is diabetic retinopathy. Diabetic retinopathy is a vascular disease of the retina and it is one of the leading causes of visual loss globally. It can be challenging for a person to notice any symptoms until he is near to lose his sight. The early detection of the disease and proper care can help to slow down the progress of diabetic retinopathy. Since microaneurysms are the first clinical symptoms of the disease, efficient detection can aid early diagnosis.

The aim of this thesis was to design and implement a computer-aided detection system for diabetic retinopathy, microaneurysms and small hemorrhages. The process pipeline included image preprocessing, extraction of 21 texture features, feature selection and image classification with a support vector machine. The system was tested on Messidor, DIARETDB1 and e-ophta MA datasets.

The first experiment studied the classification of Messidor images into healthy or diabetic retinopathy. The best average accuracy was 69 % with 72 % average sensitivity and 66 % average specificity. The global texture features were not descriptive with a whole retinal image since presenting a whole image in a such compact form will lead into a loss of information. Additionally, resizing the images that were in different sizes and combining three diabetic retinopathy severity classes into one class created variation within the final class and affected the classification negatively.

The second experiment studied the classification of DIAREDB1 and e-ophta images into microaneurysms and normal by dividing the retinal images into overlapping blocks of size 150*150. For DIARETDB1 set, the best average accuracy was 92 % with 92 % average sensitivity and 92 % average specificity with the combination of LBP and Laws features. The texture features were the most successful on the third experiment where the e-ophta images were divided into small blocks of size 50*50 with 50 % overlap. The best average accuracy was 96 % with 96 % average sensitivity and 96 % average specificity with the combination of LBP, Laws, sum average, autocorrelation, difference entropy and Gabor features. The texture features were more descriptive in local setting since then they could extract finer details.

To develop the systems further, an optic disc and image quality detectors are needed. The system is not fully automated since it needs the user to define the location of an optic disc in an image. Additionally, the image quality detector could ensure high accuracy for the detection and notify the user about low image quality.

The analysis of retinal images is laborious and subjective work. The system can ease the clinical workflow, make retinal image analysis more accurate and efficient.

REFERENCES

Abràmoff, M.D., Garvin, M.K. & Sonka, M. (2010). Retinal imaging and image analysis, IEEE reviews in biomedical engineering, Vol. 3pp. 169-208.

Acharya, U.R., Mookiah, M.R.K., Koh, J.E., Tan, J.H., Bhandary, S.V., Rao, A.K., Fujita, H., Hagiwara, Y., Chua, C.K. & Laude, A. (2016). Automated screening system for retinal health using bi-dimensional empirical mode decomposition and integrated index, Computers in biology and medicine, Vol. 75pp. 54-62.

Acharya, U.R., Ng, E.Y., Tan, J., Sree, S.V. & Ng, K. (2012). An integrated index for the identification of diabetic retinopathy stages using texture parameters, Journal of medical systems, Vol. 36(3), pp. 2011-2020.

Ahmed, A., Gibbs, P., Pickles, M. & Turnbull, L. (2013). Texture analysis in assessment and prediction of chemotherapy response in breast cancer, Journal of Magnetic Resonance Imaging, Vol. 38(1), pp. 89-101.

Antal, B. & Hajdu, A. (2012). An ensemble-based system for microaneurysm detection and diabetic retinopathy grading, IEEE transactions on biomedical engineering, Vol. 59(6), pp. 1720-1726.

Batchelor, B.G. & Whelan, P.F. (2012). Basic Machine Vision Techniques, in: Batchelor, B.G. (ed.), Machine Learning Handbook, Springer, London, UK, pp. 565-623.

British Diabetic Association (1997). Retinal photographic screening for diabetic eye disease, British Diabetic Association, London, UK, .

Chen, H. (2008). Diabetic retinopathy and associated ophthalmic disorders in: Donnelly, R. & Horton, E. (ed.), Vascular Complications of Diabetes: Current Issues in Pathogenesis and Treatment, 2nd ed., Wiley-Blackwell, Massachusetts, USA, pp. 139-150.

Decencière, E., Cazuguel, G., Zhang, X., Thibault, G., Klein, J., Meyer, F., Marcotegui, B., Quellec, G., Lamard, M. & Danno, R. (2013). TeleOphta: Machine learning and image processing methods for teleophthalmology, IRBM, Vol. 34(2), pp. 196-203.

Decencière, E., Zhang, X., Cazuguel, G., Laÿ, B., Cochener, B., Trone, C., Gain, P., Ordóñez-Varela, J., Massin, P. & Erginay, A. (2014). Feedback on a publicly distributed image database: the Messidor database, Image Analysis and Stereology, Vol. 33(3), pp. 231-234.

Devereux, R.B. (2010). Coronary artery disease and cardiomyopathy, in: Poretsky, L. (ed.), Principles of Diabetes Mellitus, 2nd ed., Springer US, Boston, MA, USA, pp. 499-513.

Dias, J.M.P., Oliveira, C.M. & da Silva Cruz, Luis A. (2014). Retinal image quality assessment using generic image quality indicators, Information Fusion, Vol. 19pp. 73-90.

Fleming, A.D., Philip, S., Goatman, K.A., Olson, J.A. & Sharp, P.F. (2006). Automated assessment of diabetic retinal image quality based on clarity and field definition, Investigative ophthalmology & visual science, Vol. 47(3), pp. 1120-1125.

Fleming, A.D., Goatman, K.A., Philip, S., Williams, G.J., Prescott, G.J., Scotland, G.S., McNamee, P., Leese, G.P., Wykes, W.N., Sharp, P.F., Olson, J.A. & Scottish Diabetic Retinopathy Clinical Research Network (2010). The role of haemorrhage and exudate detection in automated grading of diabetic retinopathy, The British journal of ophthalmology, Vol. 94(6), pp. 706-711.

Garg, S. & Davis, R.M. (2009). Diabetic retinopathy screening update, Clinical diabetes, Vol. 27(4), pp. 140-145.

Goatman, K.A., Whitwam, A.D., Manivannan, A., Olson, J.A. & Sharp, P.F. (2003). Colour normalisation of retinal images, Proceedings of Medical Image Understanding and Analysis, Sheffield, UK, July 10-11, 2013, BMVA, Sheffield, UK, pp. 49-52.

Gonzalez, R.C. & Woods, R.E. (2002). Digital image processing, 2nd ed., Prentice Hall, New Jersey, USA, 793 p.

Guyon, I. & Elisseeff, A. (2006). An introduction to feature extraction, in: Guyon, I., Gunn, S., Nikravesh, M. & Zadeh, L.A. (ed.), Feature Extraction: Foundations and Applications, 1st ed., Springer Berlin Heidelberg, Berlin, Germany, pp. 1-25.

Haghighat, M., Zonouz, S. & Abdel-Mottaleb, M. (2015). CloudID: Trustworthy cloud-based and cross-enterprise biometric identification, Expert Systems with Applications, Vol. 42(21), pp. 7905-7916.

Haralick, R.M. & Shanmugam, K. (1973). Textural features for image classification, IEEE Transactions on Systems, Man, and Cybernetics, Vol. 3(6), pp. 610-621.

Haykin, S. (1999). Neural networks: a comprehensive foundation, 2nd ed., Prentice Hall, New Jersey, USA, 340-372 p.

Hijazi, M.H.A., Coenen, F. & Zheng, Y. (2012). Data mining techniques for the screening of age-related macular degeneration, Knowledge-Based Systems, Vol. 29pp. 83-92.

Jaeger, S., Karargyris, A., Candemir, S., Folio, L., Siegelman, J., Callaghan, F., Xue, Z., Palaniappan, K., Singh, R.K. & Antani, S. (2014). Automatic tuberculosis screening using chest radiographs, IEEE Transactions on Medical Imaging, Vol. 33(2), pp. 233-245.

Kaggle 2015 Diabetic retinopathy detection, Kaggle Inc., web page. Available (accessed 17.1.2017): https://www.kaggle.com/c/diabetic-retinopathy-detection.

Kamarainen, J., Kyrki, V. & Kalviainen, H. (2006). Invariance properties of Gabor filter-based features-overview and applications, IEEE Transactions on Image Processing, Vol. 15(5), pp. 1088-1099.

Kaschke, M., Donnerhacke, K. & Rill, M.S. (2013). Optical Devices in Ophthalmology and Optometry: Technology, Design Principles and Clinical Applications, 1st ed., Wiley-VHC, Weinheim, Germany, 639 p.

Kauppi, T., Kalesnykiene, V., Kamarainen, J.-., Lensu, L., Sorri, I., Raninen, A., Voutilainen, R., Pietilä, J., Kälviäinen, H. & Uusitalo, H. (2007). DIARETDB1 diabetic retinopathy database and evaluation protocol, Proceedings of the Medical Image Understanding and Analysis, pp. 61-65.

Kauppi, T. (2010). Eye fundus image analysis for automatic detection of diabetic retinopathy, Doctoral dissertion, Lappeenranta University of Technology, 171 p. Available: http://urn.fi/URN:ISBN:978-952-265-017-7.

Kolb, H. (1995). Simple Anatomy of the Retina, in: Kolb, H., Fernandez, E. & Nelson, R. (ed.), Webvision: The Organization of the Retina and Visual System, University of Utah Health Sciences Center, Utah, USA, pp. 1-24.

Krishnan, M.M.R. & Faust, O. (2013). Automated glaucoma detection using hybrid feature extraction in retinal fundus images, Journal of Mechanics in Medicine and Biology, Vol. 13(01), pp. 1350011.

Laud, K. & Shabto, U. (2010). Diabetic Retinopathy, in: Poretsky, L. (ed.), Principles of Diabetes Mellitus, 2nd ed., Springer US, Boston, MA, USA, pp. 331-346.

Laws, K.I. (1979). Texture energy measures, Proc. Image understanding workshop, Los Angeles, CA, USA, November 7, 1979, pp. 47-51.

Leu, J.P. & Zonszein, J. (2010). Diagnostic Criteria and Classification of Diabetes, in: Poretsky, L. (ed.), Principles of Diabetes Mellitus, 2nd ed., Springer US, Boston, MA, USA, pp. 107-115.

Martini, F.H. (2006). The Special Senses, in: Fundamentals of Anatomy & Physiology, 7th ed., Pearson/Benjamin Cummings, San Francisco, USA, pp. 554-572.

MESSIDOR 2016, ADCIS, web page. Available (accessed 27.12.2016): http://www.adcis.net/en/Download-Third-Party/Messidor.html.

Meyerle, C.B., Chew, E.Y. & Ferris III, F.L. (2008). Nonproliferative diabetic retinopathy, in: Duh, E.J. (ed.), Diabetic retinopathy, Humana Press, Totowa, New Jersey, USA, pp. 3-27.

Mitchell, T.M. (1997). Machine learning, McGraw-Hill Education, New York, USA, 432 p.

Niemeijer, M., Van Ginneken, B., Cree, M.J., Mizutani, A., Quellec, G., Sánchez, C.I., Zhang, B., Hornero, R., Lamard, M. & Muramatsu, C. (2010). Retinopathy online challenge: automatic detection of microaneurysms in digital color fundus photographs, IEEE Transactions on Medical Imaging, Vol. 29(1), pp. 185-195.

Nixon, M.S. & Aguado, A.S. (2012). Feature Extraction and Image Processing for Computer Vision, 3rd ed., Elsevier Science, Missouri, USA, 628 p.

Noronha, K., Acharya, U., Nayak, K., Kamath, S. & Bhandary, S. (2012). Decision support system for diabetes retinopathy using discrete wavelet transform, Proceedings of the Institution of Mechanical Engineers, Part H: Journal of Engineering in Medicine, pp. 0954411912470240.

Odstrcilik, J., Kolar, R., Budai, A., Hornegger, J., Jan, J., Gazarek, J., Kubena, T., Cernosek, P., Svoboda, O. & Angelopoulou, E. (2013). Retinal vessel segmentation by improved matched filtering: evaluation on a new high-resolution fundus image database, IET Image Processing, Vol. 7(4), pp. 373-383.

Ojala, T., Pietikainen, M. & Maenpaa, T. (2002). Multiresolution gray-scale and rotation invariant texture classification with local binary patterns, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 24(7), pp. 971-987.

Pascolini, D. & Mariotti, S.P. (2012). Global estimates of visual impairment: 2010, The British journal of ophthalmology, Vol. 96(5), pp. 614-618.

Petkov, N. (1995). Biologically motivated computationally intensive approaches to image pattern recognition, Future Generation Computer Systems, Vol. 11(4-5), pp. 451-465.

Pietikäinen, M., Hadid, A., Zhao, G. & Ahonen, T. (2011). Computer Vision Using Local Binary Pattern, Springer, London, UK, 207 p.

Pietikäinen, M. & Zhao, G. (2009). Local texture descriptors in computer vision, Proc. of IEEE International Conference on Computer Vision, Tutorial, Kyoto, Japan, September 27, 2009, .

Pizer, S.M., Johnston, R.E., Ericksen, J.P., Yankaskas, B.C. & Muller, K.E. (1990). Contrast-limited adaptive histogram equalization: speed and effectiveness, Proceedings of the First Conference on Visualization in Biomedical Computing, Atlanta, Georgia, USA, May 22-25, 1990, IEEE Computer Society Press, pp. 337-345.

Prentašić, P., Lončarić, S., Vatavuk, Z., Benčić, G., Subašić, M., Petković, T., Dujmović, L., Malenica-Ravlić, M., Budimlija, N. & Tadić, R. (2013). Diabetic Retinopathy Image Database (DRiDB): a new database for diabetic retinopathy screening programs research, 2013 8th International Symposium on Image and Signal Processing and Analysis (ISPA), IEEE, pp. 711-716.

Reza, A.W. & Eswaran, C. (2011). A decision support system for automatic screening of non-proliferative diabetic retinopathy, Journal of medical systems, Vol. 35(1), pp. 17-24.

Riaz, F., Silva, F.B., Ribeiro, M.D. & Coimbra, M.T. (2012). Invariant Gabor texture descriptors for classification of gastroenterology images, IEEE Transactions on Biomedical Engineering, Vol. 59(10), pp. 2893-2904.

Roychowdhury, S., Koozekanani, D.D. & Parhi, K.K. (2014). Dream: Diabetic retinopathy analysis using machine learning, IEEE Journal of biomedical and health informatics, Vol. 18(5), pp. 1717-1728.

Schulze-Döbold, C., Erginay, A., Robert, N., Chabouis, A. & Massin, P. (2012). Ophdiat®: Five-year experience of a telemedical screening programme for diabetic retinopathy in Paris and the surrounding area, Diabetes & metabolism, Vol. 38(5), pp. 450-457.

Shapiro, L.G. & Stockman, G. (2001). Computer Vision, 1st ed., Prentice Hall, New Jersey, USA, 608 p.

Sikiö, M. (2016). Textural Features in Medical Magnetic Resonance Image Analysis of the Brain and Thigh Muscles, Dissertation, Tampere University of Technology. Publication Vol. 1418, 64 p. Available: http://urn.fi/URN:ISBN:978-952-15-3823-0.

Soh, L. & Tsatsoulis, C. (1999). Texture analysis of SAR sea ice imagery using gray level co-occurrence matrices, IEEE Transactions on Geoscience and Remote Sensing, Vol. 37(2), pp. 780-795.

Sonka, M., Hlavac, V. & Boyle, R. (2008). Image processing, analysis and machine vision, 3rd ed., Thomson Learning, Ontario, USA, 829 p.

Taylor, R. & Batey, D. (2012). Handbook of Retinal Screening in Diabetes: Diagnosis and Management, 2nd ed., John Wiley & Sons Inc, NJ, USA, 173 p.

Toennies, K.D. (2012). Guide to medical image analysis: methods and algorithms, Springer Science & Business Media, London, UK, 468 p.

Tuceryan, M. & Jain, A.K. (1998). Texture analysis, in: Chen, C.H., Pau, L.F. & Wang, P.S.P. (ed.), The Handbook of Pattern Recognition and Computer Vision, 2nd ed., World Scientific Publishing Co., New Jersey, USA, pp. 207-248.

Uppuluri, A. GLCM_Features4 (v1.4) - Calculates the texture features from the different GLCMs, MathWorks, web page. Available (accessed 01.08.2016): https://se.mathworks.com/matlabcentral/fileexchange/22354-glcm-features4-m--vectorized-version-of-glcm-features1-m--with-code-changes-

Widodo, A. & Yang, B. (2007). Support vector machine in machine condition monitoring and fault diagnosis, Mechanical systems and signal processing, Vol. 21(6), pp. 2560-2574.

Wilkinson, C., Ferris, F.L., Klein, R.E., Lee, P.P., Agardh, C.D., Davis, M., Dills, D., Kampik, A., Pararajasegaram, R. & Verdaguer, J.T. (2003). Proposed international clinical diabetic retinopathy and diabetic macular edema disease severity scales, Ophthalmology, Vol. 110(9), pp. 1677-1682.

World Health Organization (2016). Global report on diabetes, World Health Organization, 88 p. Available: http://www.who.int/diabetes/global-report/en/

APPENDIX A: GRAY-LEVEL CO-OCCURRENCE MATRIX RELATED STATISTICAL PARAMETER EQUATIONS

1. Angular second moment or Energy (Haralick & Shanmugam 1973)

$$f_1 = \sum_{i=1}^{A}\sum_{j=1}^{A} \boldsymbol{p}(i,j)^2$$

2. Contrast (Haralick & Shanmugam 1973)

$$f_2 = \sum_{n=0}^{A-1} n^2 \left\{\sum_{i=1}^{A}\sum_{j=1}^{A} \boldsymbol{p}(i,j) \,\middle|\, |i-j| = n\right\}$$

3. Inverse difference moment or Homogeneity (Haralick & Shanmugam 1973)

$$f_3 = \sum_{i=1}^{A}\sum_{j=1}^{A} \frac{1}{1+(i-j)^2}\boldsymbol{p}(i,j)$$

4. Entropy (Haralick & Shanmugam 1973)

$$f_4 = -\sum_{i=1}^{A}\sum_{j=1}^{A} \boldsymbol{p}(i,j)\log(\boldsymbol{p}(i,j))$$

5. Autocorrelation (Soh & Tsatsoulis 1999)

$$f_5 = \sum_{i=1}^{A}\sum_{j=1}^{A} (ij)\boldsymbol{p}(i,j)$$

6. Dissimilarity (Soh & Tsatsoulis 1999)

$$f_6 = \sum_{i=1}^{A}\sum_{j=1}^{A} |i-j| \cdot \boldsymbol{p}(i,j)$$

7. Cluster shade (Soh & Tsatsoulis 1999)

$$f_7 = \sum_{i=1}^{A}\sum_{j=1}^{A} \left(i+j-\mu_x-\mu_y\right)^3 \boldsymbol{p}(i,j)$$

8. Maximum probability (Soh & Tsatsoulis 1999)

$$f_8 = \max_{i,j} \boldsymbol{p}(i,j)$$

9. Sum average (Haralick & Shanmugam 1973)

$$f_9 = \sum_{i=2}^{2A} i p_{x+y}(i)$$

10. Sum entropy (Haralick & Shanmugam 1973)

$$f_{10} = -\sum_{i=2}^{2A} p_{x+y}(i) \log(p_{x+y}(i))$$

11. Sum variance (Haralick & Shanmugam 1973)

$$f_{11} = \sum_{i=2}^{2A} (i - f_{10})^2 p_{x+y}(i)$$

12. Difference variance (Haralick & Shanmugam 1973)

$$f_{12} = \sum_{i=0}^{A-1} \left(i - \mu_{x-y}\right)^2 p_{x-y}(i)$$

13. Difference entropy (Haralick & Shanmugam 1973)

$$f_{13} = \sum_{i=0}^{A-1} p_{x-y}(i) \log(p_{x+y}(i))$$

14. and 15. Information measures of correlation (Haralick & Shanmugam 1973)

$$f_{14} = \frac{f_4 - HXY_1}{\max(f_{4(x)}, f_{4(y)})}$$

$$f_{15} = \sqrt{1 - \exp(-2.0(HXY_2 - f_4))}$$

In the previous equations:

$$\boldsymbol{p}(i,j) = \frac{\boldsymbol{CM}(i,j)}{R_N},$$

where R_N is the amount of neighboring pixels within the area

$$p_{x+y}(k) = \left\{ \sum_{i=1}^{A} \sum_{j=1}^{A} p(i,j) \;\middle|\; \begin{matrix} i+j=k \\ k=2,3,\ldots,2A \end{matrix} \right\}$$

$$p_{x-y}(k) = \left\{ \sum_{i=1}^{A} \sum_{j=1}^{A} p(i,j) \;\middle|\; \begin{matrix} |i-j|=k \\ k=0,1,\ldots,A-1 \end{matrix} \right\}$$

$$HXY_1 = -\sum_{i=1}^{A} \sum_{j=1}^{A} p(i,j) \log(p_x(i) p_y(j))$$

$$HXY_2 = -\sum_{i=1}^{A} \sum_{j=1}^{A} p_x(i) p_y(j) \log(p_x(i) p_y(j))$$

$$\mu_x = \sum_{i=1}^{A} \sum_{j=1}^{A} i \cdot p(i,j)$$

$$\mu_y = \sum_{i=1}^{A} \sum_{j=1}^{A} j \cdot p(i,j)$$

$$\sigma_x = \sum_{i=1}^{A} \sum_{j=1}^{A} (i-\mu_x)^2 \cdot p(i,j)$$

$$\sigma_y = \sum_{i=1}^{A} \sum_{j=1}^{A} \left(j-\mu_y\right)^2 \cdot p(i,j)$$